

THESIS / THÈSE

MASTER EN INGÉNIEUR DE GESTION À FINALITÉ SPÉCIALISÉE EN DATA SCIENCE

Eléments d'analyse de la présence des marques dans l'univers digital Le cas de la communication des marques de sport sur Twitter

Bartiaux, Iseline

Award date:
2020

Awarding institution:
Université de Namur

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.



Éléments d'analyse de la présence des marques dans l'univers digital Le cas de la communication des marques de sport sur Twitter

Iseline BARTIAUX

Directeur: Prof. I. LINDEN

Mémoire présenté
en vue de l'obtention du titre de
Master 120 en ingénieur de gestion, à finalité spécialisée
en data science

ANNEE ACADEMIQUE 2019-2020

Remerciements

Pour la réalisation de mon mémoire, j'ai reçu de l'aide et des encouragements de différentes personnes que je tiens à remercier chaleureusement.

Je tiens tout particulièrement à exprimer ma reconnaissance à ma directrice de mémoire, Madame Isabelle LINDEN pour son temps investi dans un suivi régulier et des corrections éclairées mais aussi pour ses idées innovantes.

Je désire aussi montrer ma gratitude à tous mes professeurs de l'Université de Namur qui m'ont donné les outils et les connaissances nécessaires pour mettre en place les différentes parties de ce travail.

Enfin, je remercie aussi mes connaissances pour le temps passé à la relecture de cet écrit.

Table des matières

Remerciements	2
0. Introduction.....	5
1. Collecte des données.....	7
1.1. Choix du réseau social et des marques	7
1.2. Collecte des données.....	8
1.3. Nettoyage des données.....	11
1.4. Limites	14
2. Revue de la littérature.....	16
2.1. Classification de stratégies de communication.....	16
2.2. Analyse de sentiment	17
3. Analyse des données des pages	19
3.1. Première analyse descriptive	19
3.1.1. Nombre de tweets par page.....	19
3.1.2. Mots les plus récurrents.....	21
3.1.3. Nombre de retweet, favorite et followers	25
3.2. Analyse de sentiments	26
3.2.1. Principes de la reconnaissance des émotions.....	26
3.2.2. Résultats	28
3.3. Analyse du moment des publications	32
3.3.1. Semaine et évènements marquants	32
3.3.2. Jour de la semaine.....	35
3.3.3. Heure	36
3.4. Analyse de stratégie de communication	39
4. Analyse des données de la communauté.....	43
4.1. Première analyse descriptive	43
4.1.1. Analyse des mots les plus récurrents.....	43
4.1.2. Analyse du contenu des tweets populaires.....	46
4.2. Analyse de sentiments	48
4.3. Analyse du moment des publications	51
4.3.1. Semaine et évènements marquants	51
4.3.2. Jour de la semaine.....	54
4.3.3. Heure	55

5. Analyse transversale.....	56
5.1. Remarques générales.....	56
5.2. Analyse des sentiments.....	59
5.3. Analyse des moments opportuns.....	61
5.4. Stratégie de communication	62
6. Conclusion	64
Illustrations.....	66
Bibliographie.....	68
Annexes	70
ANNEXE 01 : Liste des mots à supprimer	70
ANNEXE 02 : Meilleurs moments pour poster sur Twitter pour des biens de consommation	70
ANNEXE 03 : Listes des championnats mondiaux parmi les plus grands évènements sportifs de l'année 2019	70
ANNEXE 04 : Listes comprenant les mots qui se rapportent à une stratégie de communication	72

0. Introduction

Depuis déjà quelques années, il est largement admis que les réseaux sociaux font partie intégrante de nos vies. Passer du temps sur les réseaux sociaux est devenu une activité quotidienne et incontournable pour plus de trois milliards d'utilisateurs [1]. Les entreprises ont bien compris cela et comptent bien en tirer parti. Le digital est d'ailleurs devenu la branche la plus imposante du paysage publicitaire dans la plupart des pays (GroupM, 2019). Parmi cette publicité en ligne, les réseaux sociaux représentent le deuxième segment ayant le plus grand chiffre d'affaire, après les moteurs de recherche. Cependant, il est prédit que les réseaux sociaux s'imposent comme le segment numérique le plus prospère (Statista, 2019). Cette croissance est dû au fait que les réseaux sociaux sont facilement accessibles et ils ont déjà prouvé leur efficacité. Le marketing digital et plus particulièrement le marketing sur les réseaux sociaux est donc devenu un outil incontournable pour les entreprises. Chaque entreprise a sa propre stratégie pour ce genre de promotion et chaque stratégie a son efficacité qui lui est due.

Au vu de l'affluence des différentes marques sur les réseaux sociaux et des stratégies les plus différentes les unes des autres, il nous a semblé important de se demander comment les marques utilisaient les réseaux sociaux pour faire leur publicité. Pour fixer un champ d'action à notre étude, nous avons posé des lignes directrices. Il a donc été choisi de se focaliser sur les marques de sport. De plus, en raison des politiques de confidentialité des différents réseaux sociaux, c'est Twitter qui a été choisi pour alimenter notre étude de ses données. Grâce à ces paramètres, notre question de recherche a pu se préciser pour devenir : « Comment se comportent les marques de sport sur Twitter et quel est l'effet sur leur communauté ? ».

Pour caractériser le comportement des marques, les données relatives à leur page Twitter respective seront étudiées. En ce qui concerne le comportement de leur communauté, la recherche sera faite sur les tweets comprenant le nom d'une marque. Pour répondre à notre question principale, différents éléments d'analyse ont été abordés. Le premier élément est le profil général des données et le contenu le plus récurrent apparaissant dans ces données. Une analyse de sentiments est menée en guise de deuxième élément d'analyse. Ensuite, les moments lors desquels les tweets sont postés sont comparés et finalement, une analyse de la stratégie de communication des différentes marques est présentée pour les données des pages.

Notre réflexion est construite de la façon suivante : le processus de collecte et de nettoyage de données sera tout d'abord décrit. Ensuite, une revue de la littérature présentera les travaux déjà effectués sur le même sujet que nos recherches. Les données des pages des marques seront analysées

en fonction des quatre éléments décrits ci-dessus. Par après viendra l'analyse des données des communautés avec les mêmes éléments à l'exception de la stratégie de communication qui est propre aux données des pages. Finalement, ces deux analyses seront comparées et une conclusion sera tirée.

1. Collecte des données

Dans cette première section, une présentation des choix qu'il a été nécessaire de faire pour fixer les lignes directrices de ce travail sera tout d'abord faite. Ensuite, les démarches de collecte et de nettoyage de données seront exposées. Finalement, il faut savoir qu'en travaillant avec de telles données, tout ne peut pas être réalisé comme on le souhaite. C'est pour cela que les limites de cette collecte seront explicitées.

1.1. Choix du réseau social et des marques

Le choix du réseau social s'est porté sur Twitter. Facebook avait été pris en considération en premier temps car c'est la plateforme la plus populaire en nombre d'utilisateurs [2] et beaucoup d'informations auraient pu y être extraites. Cependant, la politique de confidentialité de ce réseau social est assez stricte, ce qui empêchait de pouvoir collecter des données depuis cette plateforme. D'autres réseaux populaires ont été pris en compte comme Instagram, YouTube ou Pinterest mais ceux-ci ne répondaient pas au critère sur la forme des données. En effet, c'est du texte que nous voulons analyser et ces réseaux proposent majoritairement des images ou des vidéos. Twitter est donc un bon compromis car, même si moins populaire que Facebook, cette plateforme fait quand même partie du top 20 du nombre d'utilisateurs (WebMarketing Conseil, 2020). De plus, la politique de confidentialité est moins stricte et des packages Python sont disponibles pour collecter les données des tweets présents sur le réseau sociale. Finalement, la forme des tweets est tout à fait pertinente pour notre étude. En effet, même s'ils sont parfois agrémentés de photos, la plus grande majorité des tweets contient du texte qui peut être analysé. De plus, Twitter étant un réseau social de microblogages dont les messages sont limités à 280 caractères, les données seront plus faciles à analyser. En effet, selon Bermingham et Smeaton (2010) [3], les microblogs sont plus faciles à analyser par sentiment que les longs documents comme les blogs ou les réseaux sociaux sans limite de caractère.

Avant de commencer à recueillir les informations nécessaires à notre étude, un autre choix a dû être fait, le choix des marques. Il s'est porté sur des entreprises vendant principalement des articles de sport, que ce soit des chaussures, des vêtements ou d'autres accessoires. Ces marques doivent donc être connues majoritairement pour la vente d'articles de sport et non pour un autre type d'articles. Cependant, ces marques devaient respecter certains critères. Par exemple, la marque ne pouvait pas se concentrer sur un sport en particulier mais plutôt vendre des articles pour différents sports comme le football, la course à pied, le fitness et bien d'autres. Ensuite, il devait s'agir de marques internationales, c'est-à-dire, vendues dans les différents continents mais aussi des marques vendues

en Belgique car elles sont plus familières. Avec ces critères, le nombre de marques est encore considérable. C'est pourquoi les marques les plus répandues et connues du grand public ont été choisies. Neuf marques ont été choisies car c'est un bon compromis entre trop peu de données et trop d'informations à traiter.

Il fallait aussi que ces marques soient présentes sur Twitter car le besoin de collecter des données sur ce réseau social est primordial pour ce travail. Ces marques doivent donc avoir un compte officiel sur Twitter. Le fait que ces marques aient posté quelques tweets durant l'année 2019 est aussi primordial pour pouvoir se baser sur des tweets ayant été réalisés la même année. L'année 2019 a été choisie car c'était l'année civile complète la plus récente. La majorité des tweets doivent être en anglais car c'est une langue internationale et c'est sur celle-ci que sera faite l'analyse. En effet, se limiter au français restreindrait fortement les données à collecter. Finalement, le nom de la marque ne devait pas faire référence à autre chose que cette marque sinon les tweets récoltés auraient pu être faussés. En effet, la marque « Champion » avait été envisagée mais rechercher les tweets contenant le mot « champion » n'aurait pas été révélateur de la marque. Beaucoup d'autres tweets n'y étant pas liés auraient été insérés car « champion » peut être utilisé dans beaucoup d'autres contextes.

Les marques sélectionnées sont donc :

Adidas	Asics	New Balance
Nike	Puma	Reebok
Salomon	The North Face	Under Armour

Pour la collecte de données, seuls les tweets du compte officiel principal de la marque ont été collectés. Pour Asics, aucun compte principal n'existe, le compte relatif aux USA a donc été pris en compte car c'est le plus approvisionné en tweets et ceux-ci sont en anglais. De plus, c'est aussi le compte le plus populaire si l'on se réfère au nombre d'abonnés.

1.2. Collecte des données

Pour collecter les données de ces différentes marques sur Twitter, il fallait tout d'abord faire le choix de la technologie à utiliser. Ce choix s'est porté sur le langage de programmation Python car c'est un langage sous licence libre qui permet l'importation de différents packages en fonction des besoins. Le premier package nécessaire a été « tweepy » [4] qui est une librairie Python qui permet un accès facile à l'API de Twitter et donc de collecter des tweets pour les stocker dans une base de données

définie. Les méthodes de ce package peuvent être configurées pour collecter les informations à propos des tweets présents sur les pages et postés par les marques elles-mêmes. Grâce à ces méthodes, des informations telles que le texte, la date, la langue, le nombre de fois que le post a été retweeté ou encore le nombre de « favorites »¹ qu'à ce tweet peuvent être récoltées. Afin de stocker ces données, un autre package a dû être importé et celui-ci est « pymongo » [5] qui permet de créer et gérer une base de données MongoDB [6] en y stockant des données sous format JSON [7]. En effet, le choix du type de base de données s'est tourné vers MongoDB car il permettait de créer différentes collections pour chaque marque et ensuite, dans ces collections, chaque tweet est un document car tous les tweets ont le même format. De plus, les bases de données MongoDB stockent leurs documents sous le format JSON et ce format est adapté aux informations récoltées dans les tweets car elles sont sous forme de clé-valeur qui est le principe même du format JSON.

Deux critères de sélection ont été retenus pour la sélection des tweets. Tout d'abord, des tweets postés par les marques citées ci-dessus ont été récoltés et insérés dans la base de données MongoDB dénommée « *twitter_pages* » car ces tweets viennent des pages des différentes marques. Cette base de données comprend donc des tweets faisant référence à la stratégie de communication des marques. Les deux packages cités ci-avant ont été utilisés pour cette collecte. Ensuite, l'autre critère consiste en des tweets dans lesquels le nom d'une des marques est présent et ce, afin de savoir ce que les utilisateurs du monde entier expriment à propos de ces marques. Pour cette collecte, le package « *tweepy* » n'était pas adéquat car il ne permet pas de récolter des tweets plus vieux que vingt jours. Hors, cette collecte devait se faire sur toute la période de l'année 2019. C'est pourquoi le package « *GetOldTweets3* » [8] a été utilisé. Il n'y a pas de limite de temps avec les méthodes de ce package et il permet aussi d'obtenir les informations les plus importantes, c'est-à-dire, le texte, la date, le nombre de retweets² et le nombre de « favorites ». La langue ne fait cependant pas partie des champs disponibles avec ce package. Les données répondant à ce deuxième critère ont été stockées dans une base de données MongoDB « *twitter_communaute* ». Ce nom a été donné à la base de données car nous considérons que ces tweets sont ceux postés par la communauté de la marque, la communauté constituant un ensemble de personnes parlant de la même marque sur Twitter. Dans chacune de ces deux bases de données, neuf collections ont été créées représentant chacune les neuf marques. Les collections reprennent donc les tweets spécifiques à chaque marque. Lorsque nous parlerons de collections par après, cela signifiera donc que c'est une sous-section de la base de données réservée pour une seule marque.

¹ Le nombre de favorite est équivalent au nombre de « j'aime » déposé par les utilisateurs

² Le nombre de retweet est le nombre de fois où une publication a été partagée par un autre utilisateur

Dans la FIGURE 01, nous pouvons voir la représentation de nos deux bases de données avec, pour chacune, leurs neuf collections pour les neuf marques. Nous pouvons aussi voir des exemple de données recueillies au format JSON. Les données présentes dans la base de données « *twitter_pages* » ont été collectées avec le package « *tweepy* » et sont plus riches. En effet, plus d'informations sont disponibles pour celles-ci.

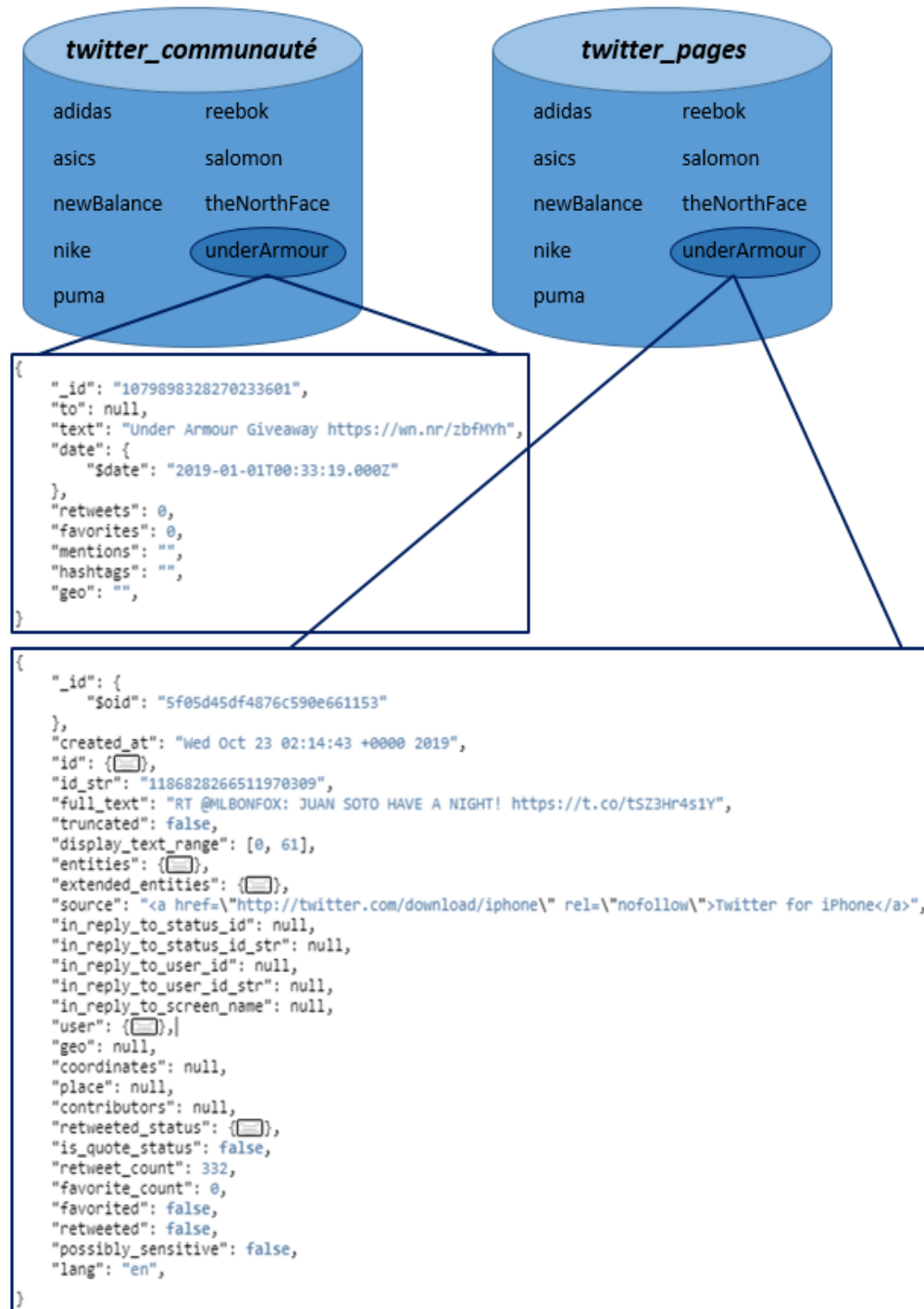


FIGURE 01 – Diagramme représentant les bases de données

1.3. Nettoyage des données

La première étape du nettoyage de données a été de sélectionner seulement les tweets qui étaient écrits en anglais. Pour cela, le package « langdetect » a été utilisé car il permet de détecter la langue d'un texte. Après cette détection, il suffisait simplement de retirer de la base de données les tweets dont la langue détectée n'était pas l'anglais. Cette étape a seulement été réalisée pour la base de données « *twitter_communauté* ». À la suite de cette étape, une moyenne de 50% des tweets ont été supprimés. Certaines collections ont perdu 1/3 de leurs tweets alors que d'autres en ont perdus presque 9/10. En effet, au départ, pour la collection « *underArmour* », 314 400 tweets ont été importés dont 223 000 étaient en anglais alors que pour la collection « *salomon* », le nombre est passé de 323 000 à 38 000. Cela pourrait suggérer que Salomon est une marque plus populaire dans les pays non-anglophones. Une autre marque ayant perdu un plus grand nombre de tweet est Adidas avec 44% restant après la suppression. La même suggestion peut être tirée pour cette dernière.

Ensuite, encore pour cette même base de données, il a fallu retirer les tweets dont le nom de la marque n'apparaissait pas dans le texte. En effet, le package « GetOldTweets3 » cité précédemment, ne faisait pas de distinction entre le fait que le nom de la marque se retrouve dans le texte du tweet ou dans un autre champ comme le nom de publicateur ou dans la description de celui-ci. La méthode ne faisait pas de différence et exportait tout, c'est pourquoi un filtre était nécessaire.

D'autres tweets qu'il a fallu supprimer sont ceux répondant aux deux critères présentés dans la section 1.2, à savoir les tweets venant de la page de la marque et incluant le nom de la marque dans son texte. Comme nous allons le découvrir dans la section 3.1.2, beaucoup de tweets postés par les marques possèdent le nom de la marque dans leur texte. Il a donc été décidé de supprimer ces tweets de la base de données « *twitter_communauté* ». En effet, ces tweets sont essentiels pour la base de données « *twitter_pages* » car ils représentent une grande partie de ce que la marque poste sur twitter. De plus, ils ne représentent pas ce que dit la communauté à propos de la marque.

Ensuite, dans la base de données « *twitter_pages* », des doublons se sont insérés à cause de boucles itératives présentes dans le processus de collecte de données. Il a donc fallu retirer des tweets pour qu'ils n'apparaissent qu'une seule fois chacun. Dans cette même base de données, il y avait aussi quelques tweets qui dataient de 2018 et qui ont été supprimés.

Finalement, un tri plus spécifique a dû être fait pour la collection « *salomon* » dans la base de données « *twitter_communauté* ». En effet, Salomon est un nom assez couramment utilisé et des personnes plus ou moins célèbres le porte comme Salomon, le personnage dans la Bible ou les footballeurs Salomon Kalou et Salomon Rondon ou encore le peintre du XVII^{ème} siècle, Salomon

Koninck et encore plein d'autres. Il y a aussi un Hôtel Salomon et les Iles Salomon. Plus de 12 000 tweets ont été supprimés à la suite de cette étape. Il est possible que certains tweets concernent encore d'autres personnes se prénommant Salomon et étant moins célèbre soient toujours présents dans la base de données mais il n'était pas possible de passer en revue les 26 000 tweets restant. La plus grosse partie a tout de même été contrôlée.

Dans la FIGURE 02, on peut voir le nombre de tweets qui ont été récoltés à la base du travail et le nombre de tweets après le nettoyage des données. C'est sur ces derniers tweets que les analyses qui suivent seront menées.

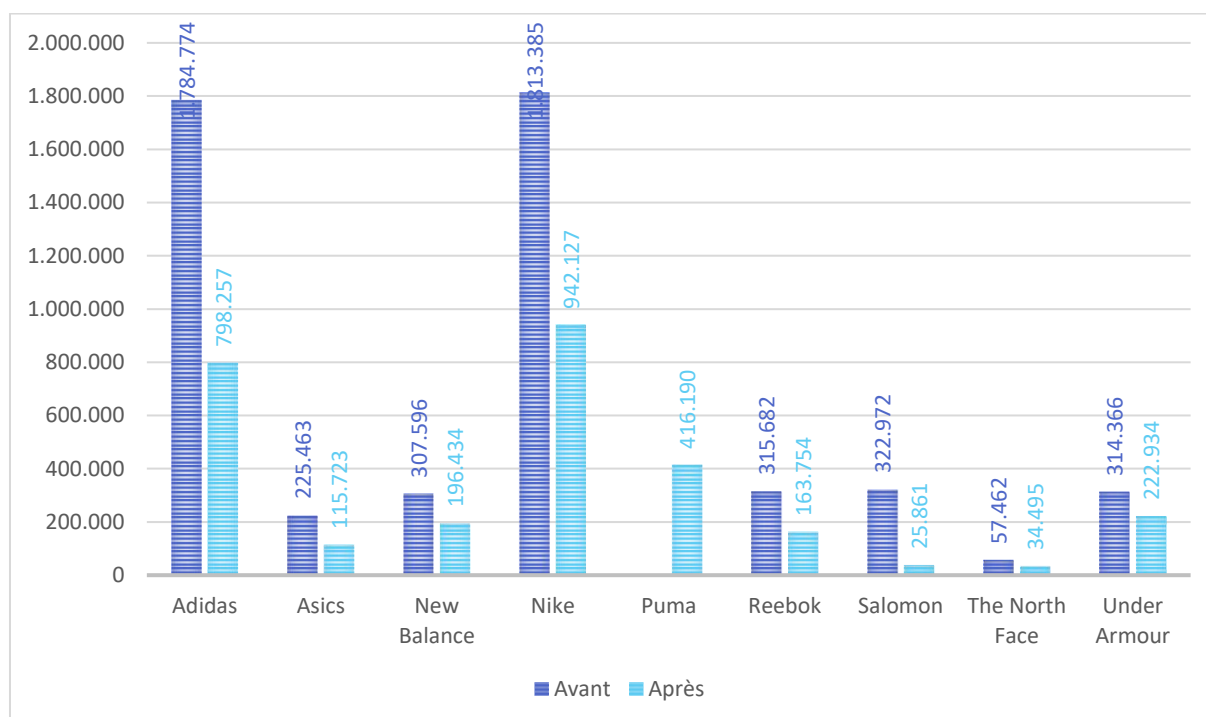


FIGURE 02 – Nombre de tweets par marques dans « *twitter_communaute* » avant et après le nettoyage des données³

Grâce à ce tableau, on peut comprendre que le nettoyage de données est une étape clé pour la réalisation d'analyses. En effet, pour une majorité des marques (**Adidas, Asics, Nike, Reebok, Salomon** et **The North Face**), entre 40% (pour **The North Face**) et 92% (pour **Salomon**) des tweets collectés en première étape n'est pas pertinente pour nos analyses. Ces tweets auraient donc biaisé toute notre étude s'ils n'avaient pas été retirés.

Le but de ce nettoyage de données était aussi de construire des listes comprenant les mots importants présents dans les tweets. Pour cela, il a tout d'abord fallu supprimer les URLs présents dans les tweets. En effet, les tweets contiennent souvent des URLs qui renvoient à des sites Internet,

³ Le nombre de tweets avant nettoyage n'est malheureusement pas disponible pour Puma car les tweets ont été supprimés avant la collecte de cette information

cependant, les mots contenus dans ces URLs ne sont pas pertinents pour l'analyse future. De plus, ces liens contiennent des signes de ponctuation qui pourraient fausser les étapes suivantes. Pour supprimer ces URLs, la fonction « `remove_url` » a été empruntée du site Earth Lab et utilisée sur toutes les données. La seconde étape était de séparer les textes en fonction de certains éléments comme les espaces, les virgules ou tout autre élément de ponctuation. Cela a été fait grâce à la méthode `word_tokenize` du package « `nlk.tokenize` » (Natural Language ToolKit) [9] qui renvoie une copie décomposée d'une chaîne de caractères, en utilisant le tokenizer de mots recommandé par NLTK. Après cette étape, il a été nécessaire de supprimer tous les éléments de ponctuation car même si ils ont servis à séparer les textes, ils n'ont pas pour autant été supprimés. Ensuite, tous les mots présents dans les listes ont été transformés pour ne contenir que des lettres minuscules, ce qui allait faciliter les étapes de comparaison suivantes. L'étape suivante a été de supprimer les mots qui n'apportaient pas de valeur ajoutée. En effet, des mots comme « un », « sur », « le », « et » sont abondamment présents dans les textes mais n'apportent pas d'indications sur le contenu des textes. Pour cela, la liste « `stopwords` », voir ANNEXE 01, du package « `nlk.corpus` » a été mise à profit et ce, en supprimant, dans les listes récemment créées comprenant tous les mots présents dans les textes, les mots présents dans cette liste.

Le tweet

“Still hopeful salomon @salomonrunning will make my #newyear2019 special by introducing #lightup #salomonsneakers. #stillhopeful #ibelieve in #salomonsports #bluegreen lightup #xtours!
https://instagram.com/p/BsFOWRvHCgc/?utm_source=ig_twitter_share&igshid=1lqc9n0rvfdo8...” @Nadia_Salomon

sera donc transformé en la liste:

```
["still", "hopeful", "salomon", "salomonrunning", "make", "special", "introducing",  
"lightup", "salomonsneakers", "stillhopeful", "ibelieve", "salomonsports",  
"bluegreen", "lightup", "xtours"]
```

La FIGURE 03 présente un résumé sous forme de diagramme du processus de collecte de données et de nettoyage qui vient d'être explicité en détail.

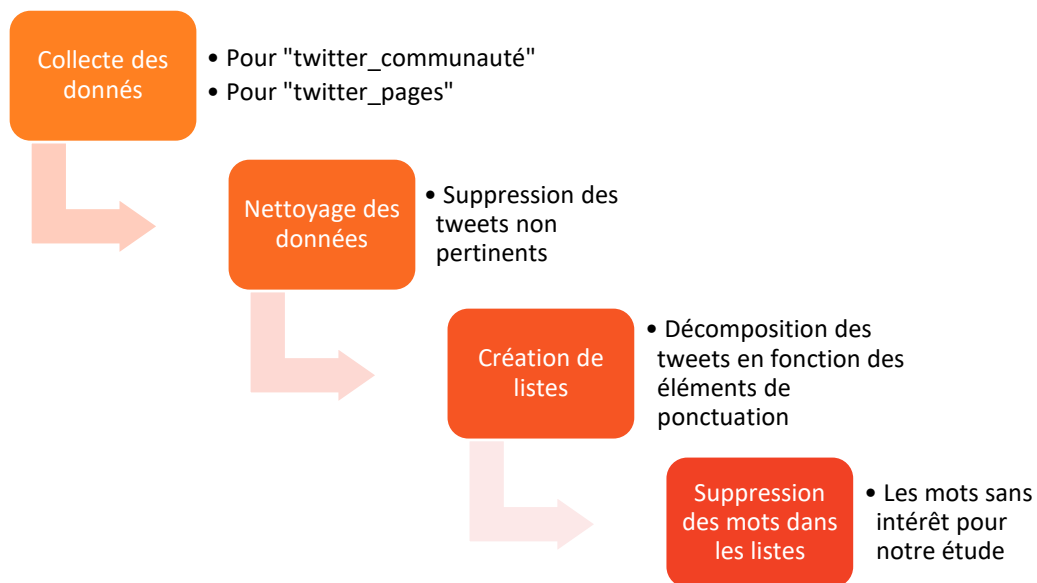


FIGURE 03 – Diagramme représentant le processus de collecte et de nettoyage des données

1.4. Limites

Comme expliqué précédemment, il n'est pas possible de faire tout ce que l'on souhaite lorsqu'on travaille avec les données d'un réseau social. Pour cette raison, certaines barrières se sont dressées lors de la collecte de données et vont être décrites ici pour affirmer de la transparence de notre étude.

Comme expliqué dans la sous-section précédente, les tweets n'étant pas en anglais ont été supprimés de la base de données « *twitter_communauté* ». Cependant, la deuxième base de données « *twitter_pages* » contenait aussi ce genre de tweets mais n'ont pas été supprimés. En effet, ces tweets étaient nécessaires pour certaines analyses comme pour le calcul du nombre de tweets ou du nombre de « favorite » moyen. Le contenu de ceux-ci ne sera cependant pas étudié à l'exception des mots populaires écrits en anglais qui pourraient s'y retrouver.

Dans les tweets collectés pour la base de données « *twitter_communauté* », certains peuvent se rapporter à autre chose que la marque. En effet, comme expliqué dans la sous-section précédente, le mot Salomon peut référer à un tas d'autres choses que la marque de sport. Pour ce cas-ci, le problème a été réglé car il était possible de supprimer les tweets contenant le nom des personnages se prénommant Salomon. Cependant, pour **Puma** par exemple, le mot peut aussi faire référence à l'animal et il n'y avait pas moyen de trier ce genre de tweets ni de passer en revue toute la collection Puma qui contient plus de 40.000 tweets. L'hypothèse a donc été faite que le nombre de tweet parlant de la marque surpasserait celui des tweets parlant de l'animal. D'ailleurs, lorsqu'on tape « puma » dans la barre de recherche de Twitter, les tweets affichés sont en relation avec la marque. Pour confirmer cette hypothèse, les cent tweets ayant le plus de favorite ont été lus et parmi ceux-ci, seulement deux parlaient de l'animal. Si l'on fait l'hypothèse que ces cent tweets sont représentatifs de la collection, il

n'y aurait que 2% des tweets qui ne sont pas pertinents. Ce pourcentage étant faible, ces tweets n'auront un impact que très minime sur notre étude.

De plus, toujours pour la même base de données, tous les tweets n'ont pas pu être récoltés pour toutes les marques (**Nike** et **Adidas**). En effet, l'algorithme utilisé ne permettait de collecter qu'un nombre de tweets limités par jour. Pour la majorité des marques, cette limite était largement suffisante mais pour les marques les plus connues, ça ne l'était pas. Cependant, au vu du nombre total de tweets récoltés pour ces deux marques, il y aura assez de tweets pour faire nos analyses.

Dans la base de données « *twitter_pages* », la collecte des tweets ne s'est pas faite correctement dans un premier temps. Les textes des tweets les plus longs n'ont pas été collectés dans leur intégralité. En effet, après un certain nombre de caractères, la ponctuation « ... » apparaissait pour remplacer la fin du texte. Ce problème a été résolu pour la majorité des marques. Cependant, pour **Asics** et **Puma**, certains tweets originaux ont été supprimés de Twitter entre temps et n'ont donc pas pu être retrouvés. Pour ces deux marques, il manquera donc certains mots dans les listes créées.

2. Revue de la littérature

Maintenant que nous savons de quoi se composent nos données, nous pouvons commencer à les analyser. Avant cela, une recherche dans la littérature est nécessaire pour comprendre les différentes classifications de stratégie de communication des entreprises mais aussi pour s'imprégner de travaux déjà réalisés sur l'analyse de sentiment.

2.1. Classification de stratégies de communication

Dans la littérature, il y a eu peu de tentatives pour élaborer une classification de stratégies de communication. Certains modèles existent tout de même mais ne sont pas très populaires. Le modèle à six segments de Taylor (1999) [10] est un des premiers à apparaître. Par la suite, Golan et Zaidner (2008) [11] utiliseront le modèle de Taylor pour l'appliquer à une analyse de la publicité virale. Dans leur publication, Golan et Zaidner redéfinissent ce modèle qui est composé de deux approches, la vue de transmission, similaire à une approche informationnelle et la vue rituelle similaire à une approche transformationnelle. Ces deux vues sont elles-mêmes divisées en trois segments : ration, besoin aigu et routine pour la vue de transmission et ego, social et sensoriel pour la vue rituelle. Ces segments ne sont cependant pas adaptés à une communication sur les réseaux sociaux et la détection de ceux-ci ne seraient pas aisément automatisable.

L'article de Jensen et Jepsen (2008) [12] s'intéresse quant à lui aux outils de communication pour le marketing en ligne. Ils tentent de catégoriser ces outils en différentes disciplines et les catégories qui émergent sont les suivantes : publicité, relations publiques, promotion des ventes, marketing direct et communication personnelles. Ces groupes sont plus adaptés aux réseaux sociaux mais leur granularité n'est pas assez fine pour pouvoir mener des analyses.

Kim et Ko (2012) [13] traitent aussi de stratégie de marketing en ligne et plus précisément de marketing des réseaux sociaux. Ils distinguent cinq types d'activités : le divertissement, l'interaction, la tendance, la customisation et le bouche-à-oreille. Ces activités sont cependant spécifiques aux marques de luxe et ne pourront donc pas être utilisées dans le cadre de cette analyse.

Davis, Piven et Breatzle (2014) [14] ont quant à eux construit le modèle à cinq sources. Ce modèle identifie cinq moteurs principaux de la consommation de marque dans une communauté de réseaux sociaux : fonctionnel, émotionnel, auto-orienté, social et relationnel. Tout comme pour Jensen et Jepsen (2008), le niveau de granularité de la catégorisation n'est pas assez fin.

Face au peu d'études concernant des classifications pour la communication sur les réseaux sociaux, Coursaris, Van Osch et Balogh (2013) [15] ont décidé de mener une étude longitudinale de la théorie

multi-terrains pour créer une telle typologie. Pour cela, ils ont analysé les publications de six semaines sur Facebook de trois entreprises notables. La typologie qu'ils proposent englobe sept catégories : la **notoriété de la marque**, la **responsabilité sociale des entreprises**, le **service clients**, l'**engagement**, la **notoriété des produits**, les **promotions** et la **saisonnalité**. Voici les définitions qu'ils affectent à chacune de ces catégories :

- **Notoriété de la marque** : publications qui renforcent la présence et l'attractivité des entreprises sur le marché des consommateurs numériques
- **Responsabilité sociale des entreprises** : publications qui construisent une image de marque de la participation au soutien et au renforcement de la communauté, principalement auprès des consommateurs socialement conscients
- **Service clients** : publications qui renforcent la connaissance des consommateurs sur les changements de produits, d'industrie et de marque
- **Engagement** : publications qui construisent des liens avec les consommateurs et des communautés via des interactions directes avec la marque
- **Notoriété des produits** : toutes les publications qui renforcent la connaissance, la compréhension et l'existence de produits
- **Promotion** : publications qui visent à simuler immédiatement ou dans un futur proche par des incitations monétaires
- **Saisonnalité** : publications qui rappellent et informent les consommateurs sur des événements saisonniers et annuels et des produits connexes de la marque

Cette typologie semble complète et d'un bon niveau de granularité. De plus, elle s'adapte facilement à nos données et des techniques peuvent être mises en place pour automatiser la détection de ces différentes stratégies. Dans ce mémoire, nous allons l'appliquer à un autre réseau social, à savoir Twitter.

En ce qui concerne la stratégie de fréquence, Ashley et Tuten (2015) [16] affirment qu'il y a un taux de fréquence minimal à respecter pour garantir une motivation de la part des utilisateurs. Les marques doivent faire des réactualisations et donner des incitations à la participation fréquemment.

2.2. Analyse de sentiment

Des travaux sur l'analyse de sentiments sont facilement trouvables dans les articles scientifiques. Ce sont généralement des études portant sur la classification d'un texte en valeur positive, négative ou neutre. Par exemple, Troussas, Virvou et al. (2013) [17] ont réalisé une reconnaissance automatique des trois émotions de base dans des textes en ligne. Ils ont utilisé la méthode de machine learning, Naïve Bayes Classifier pour exécuter la tâche. En effet, selon Boiy, Hens et al. (2007) [18], les méthodes

de machine learning sont celles qui ont le plus de précision en ce qui concerne la détection de sentiments dans des textes. Medhat, Hassan et Korashy comparent les différents algorithmes d'apprentissage automatique, à savoir les classificateurs probabilistes, les classificateurs Naïve Bayes, le réseau Bayésien, les classificateurs à entropie maximale, le classificateur « Support Vector Machine » (SVM), le « Neural Network », les arbres de décision et les classificateurs basés sur les règles. Naïve Bayes et Support Vector Machine sont les algorithmes les plus utilisés pour résoudre des problèmes de reconnaissance de sentiments. Ce sont des modèles de référence. Le premier, le modèle de classification Naïve Bayes, est le classificateur le plus simple et le plus couramment utilisé. Il calcule la probabilité conditionnelle d'une classe, en fonction de la distribution des mots dans le document. Le deuxième, le SVM, est parfaitement adapté pour les données textuelles en raison de la nature clairsemée du texte.

Le choix entre les deux modèles s'est fait sur les conseils du site scikit-learn [19], un site proposant une bibliothèque libre Python destinée à l'apprentissage automatique. Ce site est largement utilisé dans le monde du machine learning et est donc fiable. Il nous conseille d'utiliser une variante de Naïve Bayes s'appelant MultinomialNB. C'est donc cet algorithme que nous utiliserons dans ce travail.

Dans notre cas, nous avons voulu faire une analyse plus précise que simplement différencier les sentiments positifs des sentiments négatifs. Nous voulions repérer plusieurs sentiments différents comme la joie, l'amour, la colère, la tristesse, la peur, la surprise et le sentiment neutre. Chaffar et Inkpen (2011) [20] ont utilisé l'algorithme « Support Vector Machine » pour reconnaître six émotions dans des textes. Pour ce faire, ils ont entraîné leur modèle avec des phrases venant de titres de l'actualité, de contes de fées et de blogs. Nous entraînerons donc notre algorithme pour qu'il puisse retrouver autant de sentiments différents dans nos données.

3. Analyse des données des pages

Après avoir déterminé notre technique pour collecter et nettoyer les données et avoir défini les concepts nécessaires pour nos analyses, nous pouvons dorénavant commencer à analyser ces données sous différentes coutures. Nous allons tout d’abord nous concentrer sur la communication que les différentes marques ont choisi d’adopter. La base de données « *twitter_pages* » sera utilisée pour la première partie de notre étude. Une première analyse descriptive sera menée pour comprendre un peu mieux cette base de données. Ensuite, une analyse de sentiment avec trois ensembles d’entraînement différents sera réalisée. Une analyse des moments lors desquels les tweets sont postés sera aussi faite et finalement, nous classifions les tweets en fonction des stratégies trouvées dans la section précédente.

Dans cette section, nous basons nos analyses sur la première base de données nettoyée intitulée « *twitter_pages* » et dont le nombre de tweets varie entre 100 et 3000 par marque pour l’année 2019, ce qui fait un total d’environ 8500 tweets.

3.1. Première analyse descriptive

3.1.1. Nombre de tweets par page

L’analyse la plus basique mais qui apporte tout de même des informations importantes est celle reprenant le nombre de tweets collectés par page. La FIGURE 04 nous présente ces résultats.

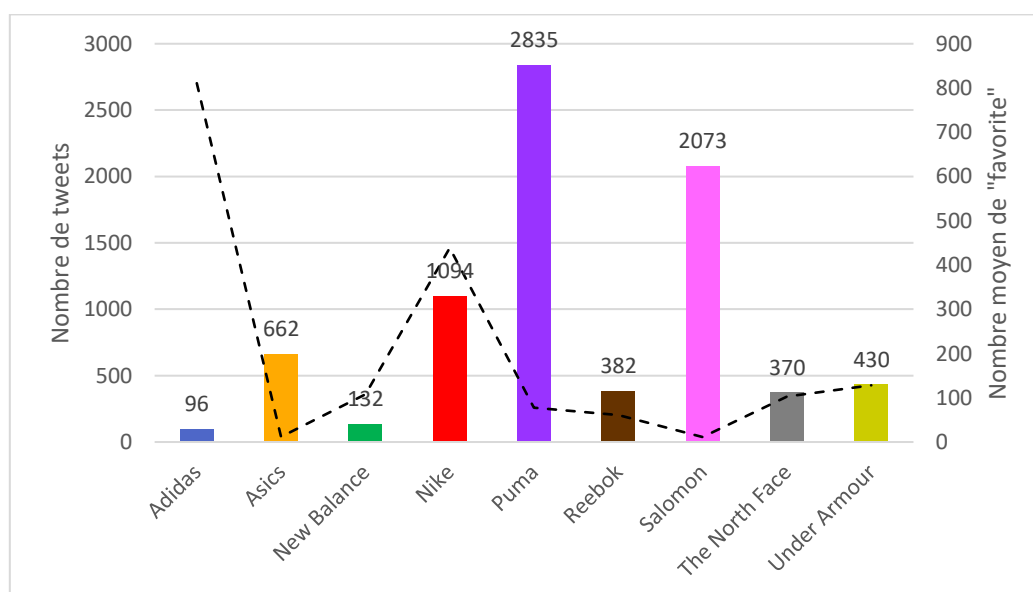


FIGURE 04 – Nombre de tweets et de « favorite » par page des marques

En 2019, avec 2835 tweets, **Puma** a posté le plus de tweets parmi nos neuf marques, suivi par **Salomon** (2073) et ensuite **Nike** (1094). Puma a donc une moyenne d'environ huit tweets par jour. **Asics** (662), **Reebok** (382), **The North Face** (370) et **Under Armour** (430) sont loin derrière Puma avec une moyenne de un ou deux tweets par jour tandis qu'**Adidas** (96) et **New Balance** (132) ne suivent pas la même tendance avec seulement environ un tweet tous les trois jours.

Cette première analyse peut nous donner une idée sur les stratégies de fréquence qu'utilisent les marques. Au vu de ce graphique, **Puma**, **Salomon** et **Nike** suivent les conseils émis par Ashley et Tuten (2015) [16] affirmant qu'il faut avoir un certain niveau de fréquence dans ses diverses actions marketing pour s'assurer de garder la motivation des utilisateurs. A l'inverse, **Adidas** et **New Balance** n'ont pas suivi ce conseil à la lettre. Cependant, au vu du nombre moyen de « favorite » et de la notoriété d'Adidas, il est peu probable que sa stratégie soit inefficace. Le fait de poster un tweet tous les trois jours doit donc être suffisant pour garder l'attention de ses utilisateurs.

Le nombre de tweets n'indique donc pas la grandeur de la marque, ni l'efficacité de leur stratégie de communication mais ils donnent une bonne idée de la fréquence à laquelle la marque poste sur Twitter. En effet, une marque de la taille d'**Adidas** n'est certainement pas plus petite que **Salomon** et encore moins dans les proportions indiquées par ces chiffres. De plus, tant que le minimum est respecté, ce n'est pas parce qu'une marque poste plus de publications que la communication est plus efficace. Le contenu de la plupart des publications pour **Puma**, **Salomon** ou **Nike** est indéniablement un contenu spécifique à certains utilisateurs, comme des réponses à des questions posées par un utilisateur.

Nous pouvons effectivement remarquer avec le nombre moyen de « favorite » sur la FIGURE 04 que ce n'est pas parce qu'une marque est plus présente sur Twitter qu'elle obtiendra plus souvent des réactions positives à ses publications. C'est d'ailleurs plutôt l'inverse. En effet, lorsqu'une marque publie beaucoup de tweets, ce sont certainement beaucoup de réponses à des utilisateurs et ce sont des tweets qui ne reçoivent pas beaucoup de « favorite ». Cela ne veut pas dire pour autant que c'est une mauvaise stratégie. En effet, l'utilisateur qui aura reçu une réponse sera peut-être satisfait que la marque prête attention à son problème ou à son exploit. Pour illustrer cela, la FIGURE 05 nous montre le nombre total de tweets par marque en comparaison avec le nombre de tweets lorsque tous ceux qui sont des réponses ont été supprimés. Une réponse signifie que c'est une publication créée pour répondre à un utilisateur et non une publication à but purement marketing. De plus, le nombre moyen de favorite a été ajouté et il a été calculé sur les deux totaux cités juste avant.

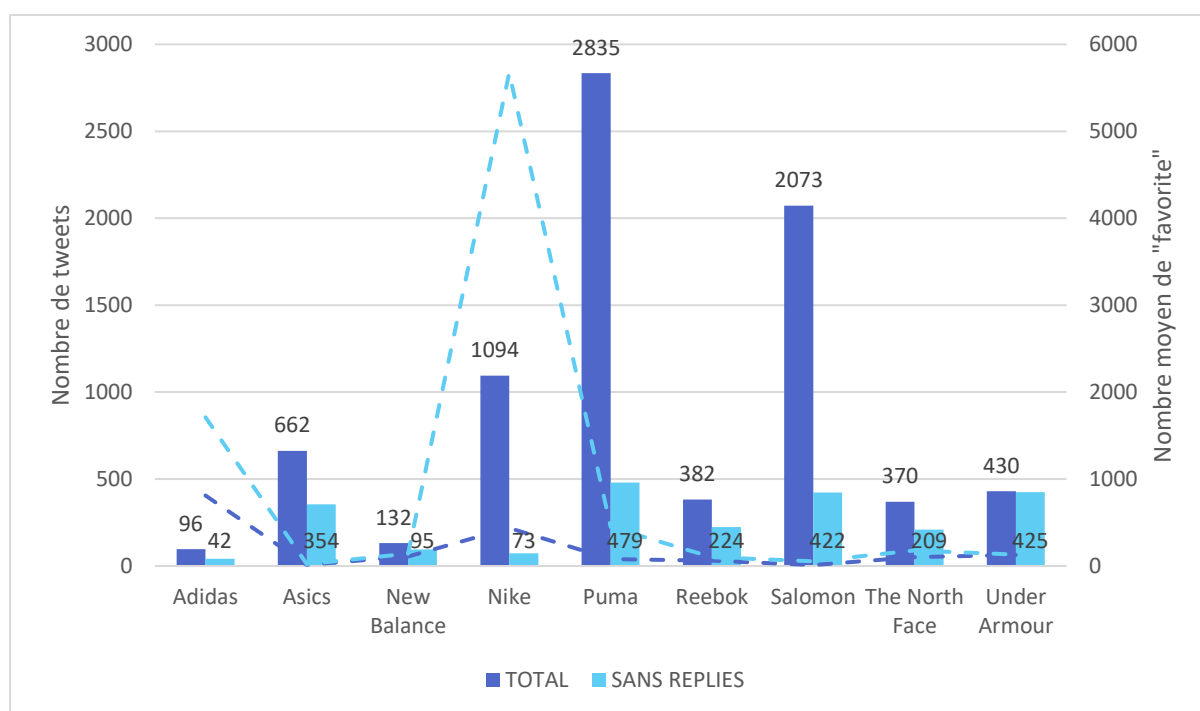


FIGURE 05 – Nombre de tweets et de « favorite » par marque pour tous les tweets et pour les tweets sans les réponses

Dans cette figure, nous pouvons effectivement constater que la majorité des tweets postés par les marques sont en réalité des réponses. Cette proportion est d'autant plus grande lorsque c'est une marque qui publie beaucoup comme **Nike** ou **Puma**. Au final, lorsque les réponses sont supprimées, le nombre de tweets est assez similaire pour toutes les marques car il varie entre 42 (pour **Adidas**) et 479 (pour **Puma**). La différence entre les stratégies de fréquence est donc moindre et on peut comprendre que même Adidas respecte le taux de fréquence minimal au vu du taux maximal qui n'est pas nettement supérieur. Finalement, le nombre moyen de « favorite » est généralement plus élevé lorsque les réponses ont été enlevées. Ceci signifie que les réponses ne rapportent pas beaucoup de mentions « favorite ». Cette différence est assez frappante pour **Nike** dont le taux de tweets après suppression n'atteint que 6.7% du total des tweets. Cela implique que les 73 tweets restant sont bien plus populaire au niveau des mentions « favorite » que les 1021 qui ont été supprimés.

3.1.2. Mots les plus récurrents

Pour comprendre globalement de quoi traitent les tweets de notre première base de données, nous avons relevé les mots les plus fréquents pour chaque marque. Pour ce faire, les mots des listes créées en section 1.3 ont été rassemblés et la fréquence de chacun des mots a été calculée. Le TABLEAU 01 présente ces résultats en reprenant les vingt mots avec la fréquence la plus haute.

Dans ce tableau, on peut tout d'abord constater sur quel sport ces marques mettent l'accent. En effet, **Adidas**, **Asics** et **Salomon** se concentrent plutôt sur le « *running* » tandis que **Reebok** se

concentre sur le « *crossfit* ». Pour **New Balance**, c'est le football avec « *nbfootball* ». De plus, les trois lettres « *lfc* » qui apparaissent pour New Balance sont l'abréviation pour Liverpool Football Club. Le mot « *reds* » fait aussi référence à cette équipe, ce qui peut laisser suggérer que New Balance articule une grande partie de sa communication sur le sponsoring avec le club anglais. New Balance est en effet équipementier du FC Liverpool. Le futur de leur communication pourrait être intéressant à étudier car Nike a repris le contrat avec le club de football, New Balance va donc devoir trouver d'autres sujets à communiquer.

Certains noms de sportifs célèbres et un peu moins célèbres apparaissent également pour d'autres marques. « *Coco Gauff* », la jeune joueuse de tennis pour **New Balance**, « *JJ Watt* », le joueur de football américain pour **Reebok**, le moins célèbre coureur « *Andre Cosen* » pour **Salomon** et finalement le basketteur de NBA « *Stephen Curry* » pour **Under Armour**. Tous ces sportifs ont un contrat avec la marque qui leur est associée, certains ont même des articles portant leur nom dans les collections des marques. Les collections sont aussi présentes dans certaines colonnes du tableau. « *Futurelight* » pour **The North Face**, « *UA Hovr* », « *UA Rush* » et « *Curry6* » pour **Under Armour**.

Le nom de la marque se trouve à chaque fois dans les vingt mots les plus récurrents. La position que ce mot occupe dans le tableau veut peut-être signifier quelque chose. **Nike** ayant son nom apparaissant dans le haut du tableau, cela pourrait signifier que c'est une marque qui a choisi d'axer sa stratégie sur le fait qu'elle possède déjà une grande notoriété et elle joue donc de cette notoriété pour attirer les consommateurs.

Asics, Puma, Reebok, Salomon et The North Face ont tous le mot « *sorry* » dans la liste des mots les plus fréquents. Le mot n'est pas dans la liste pour **New Balance** mais il apparaît en 30^{ème} position pour la marque qu'on peut donc considérer à l'identique des autres. Cela peut signifier que ces marques ont besoin de s'excuser pour créer une relation avec la clientèle. Chose que **Nike** ou **Adidas** ne doit pas faire car ils bénéficient d'une popularité plus importante.

Adidas, New Balance, Puma et Under Armour ont la chaîne de caractère « *rt* » dans les mots les plus fréquents. « *rt* » correspond à retweet et signifie que la marque n'est pas réellement l'auteur du tweet, elle repartage simplement un tweet qui a été écrit par une autre personne. Les retweets sont souvent faits quand une personne célèbre poste quelque chose de relatif à la marque ou quand un client poste sa satisfaction ou enfin quand une affirmation est dans la même optique que la marque et qu'elle veut donc la partager.

Le dernier mot de la liste de **Reebok** est « *crossfitgames* » et fait référence à la compétition annuelle de CrossFit parrainée par la marque. Comme pour New Balance et le club de foot de Liverpool, on peut penser que Reebok a choisi d'axer une partie de sa communication sur cet événement. Cet

évènement est très populaire dans l'univers du CrossFit. Comme nous avons vu juste avant que Reebok axait sa communication sur ce sport, c'est donc un bon choix stratégique pour la marque de sponsoriser cet évènement.

Le mot « *salomonwmn* » demande un peu d'attention aussi. Il réfère à une campagne de la marque **Salomon** qui a pour but de changer les fausses idées reçues à propos des femmes et de les encourager à se mettre au sport. C'est, selon le directeur général de Salomon, la campagne la plus populaire qu'ils n'ont jamais créée. De plus, un concours portant ce nom a été lancé pour gagner certains articles de la marque mais aussi une participation à un week-end multisport exclusivement féminin. La marque a aussi le mot « *salomontv* » dans sa liste. C'est le nom de la chaîne YouTube de la marque et un site Internet contenant des vidéos d'exploits sportifs porte aussi ce nom. La marque aime donc partager les expériences de ses athlètes avec ses utilisateurs.

Les slogans des différentes marques apparaissent aussi dans les mots les plus fréquents. Pour **Under Armour**, c'est même le premier, « *We Will* ». Il n'est pas loin derrière pour **The North Face** car il est en quatrième position, « *Never Stop Exploring* ». Il est en dixième position pour **Salomon** avec « *Time To Play* ». Pour d'autres marques comme **New Balance**, **Nike** et son célèbre slogan « *Just Do It* », **Reebok** et **Puma**, le slogan n'apparaît pas dans notre tableau mais ils sont tout de même présents en 42^{ème}, 83^{ème}, 166^{ème} et 387^{ème} position respectivement. D'autres associations de mots liées aux marques apparaissent aussi comme « *Here to create* » pour **Adidas**, « *Win the long run* » pour **Asics** et « *We got now* » pour **New Balance**. Nous pouvons donc ici déceler quelles entreprises choisissent de renforcer leurs images de marques en faisant apparaître leur slogan dans leur communication.

Comme nous pouvons le voir ici, avec une simple analyse des mots les plus fréquents, nous apprenons déjà beaucoup sur les sujets autour desquels les marques choisissent d'axer leur communication. Il sera intéressant de croiser cette analyse avec les autres pour voir si ces sujets ressortent de la même manière.

	Adidas		Asics		New Balance		Nike		Puma		Reebok		Salomon		The North Face		Under Armour	
	words	count	words	count	words	count	words	count	words	count	words	count	words	count	words	count	words	count
1	creatorsunit e	23	us	187	rt	47	nike	167	us	1896	us	78	thanks	498	collection	90	wewill	108
2	hear	22	dm	128	nbfootball	26	us	138	hey	1057	new	58	hi	406	us	85	shop	97
3	us	21	hear	113	nbhoops	24	hey	107	please	1032	available	55	contact	288	warranty	74	new	67
4	conversation	19	would	105	newbalance	19	available	102	contact	749	reebok	54	us	283	neverstope xploring	59	rt	52
5	first	17	sorry	85	lfc	13	air	102	help	704	dm	37	please	282	please	57	uahovr	49
6	adidas	16	running	81	team	13	check	92	drop	635	may	37	team	214	full	52	stephencu rry30	47
7	thank	16	mind	75	cocogauff	11	keep	91	hi	576	please	36	hear	191	photo	50	get	44
8	amp	14	shoe	73	making	10	de	88	sorry	448	amp	33	sorry	179	form	50	collection	40
9	ideas	14	run	71	new	10	site	79	hear	431	like	33	see	163	see	43	ua	39
10	share	13	winthelong run	69	world	9	know	76	messag e	430	hi	31	timetopl ay	151	sorry	42	latest	37
11	heretocreate	13	get	63	reds	8	un	72	dm	421	love	29	help	138	new	33	uarush	35
12	joining	13	thanks	52	see	8	time	69	assist	383	order	29	look	133	futurelight	31	game	34
13	love	12	sending	49	championshi p	7	let	69	check	332	next	28	awesome	125	would	28	shoe	28
14	running	12	feedback	45	usatfoutdoor s	7	updates	66	hello	307	energy	27	anzece se n	120	link	27	training	28
15	feedback	12	new	43	intern	7	que	63	send	235	crossfit	27	like	103	hi	26	ready	26
16	say	11	help	42	bazleydarius	7	like	62	rt	224	sorry	25	new	93	climbing	26	teamua	24
17	change	11	like	42	congrats	6	te	60	note	194	know	23	get	87	series	24	curry6	24
18	rt	10	bottom	40	im	6	get	59	puma	178	jjwatt	22	take	83	hear	24	available	22
19	wed	10	asics	39	rounds	6	got	58	look	135	apologize	21	salomonr unning	83	happened	24	performan ce	22
20	creatorsunit e	5	team	39	wegotnow	5	en	58	update s	130	crossfitga mes	21	salomon wmn	82	departmen t	24	work	22

TABLEAU 01 – Mots les plus récurrents et leur nombre d'occurrence par marque pour la base de données « *twitter_pages* »

3.1.3. Nombre de retweet, favorite et followers

Pour se faire une idée de la popularité des différentes marques sur Twitter, il peut être intéressant d'analyser le nombre de *retweet* de *favorite* qu'obtiennent ces marques sur leurs publications. De manière plus générale, le nombre d'*abonnés (followers)*⁴ sur la page des marques peut aussi donner un indice sur la popularité de la marque sur le réseau social. Le TABLEAU 02 nous présente ces résultats. Le nombre d'abonnés a été relevé le 23/07/2020 et a pu évoluer depuis lors.

Marque	Maximum retweet	Moyenne retweet	Maximum favorite	Moyenne favorite	Nombre de followers
Adidas	4963	265,68	17.153	790,16	3.788.472
Asics	178	2,48	736	10,98	96.725 ⁵
New Balance	12.005	342,99	2836	108,26	256.051
Nike	95.401	283,64	276.177	958,16	8.208.439
Puma	26.413	66,27	67.288	76,71	1.647.478
Reebok	5836	26,59	4508	70,16	716.095
Salomon	177	1,8	1193	11,28	79.602
The North Face	1044	19,17	6547	103,3	477.769
Under Armour	167.015	489,43	3232	129,36	960.905

TABLEAU 02 - Statistiques descriptives du nombre de retweet et de favorite

L'analyse du nombre de retweets et de favorites présents dans le TABLEAU 02, démontre qu'en moyenne, sur leurs pages, c'est **Nike** (790) qui a le plus de *favorites* et c'est **Under Armour** (489) qui a le plus de *retweets*. Au contraire, c'est **Asics** (2 et 11) et **Salomon** (2 et 11) qui ont le moins de *retweets* et de *favorites* sur les tweets qu'ils ont postés. Ceci peut être expliqué par le fait que ce sont les pages ayant le moins d'*abonnés* et donc moins de personnes sont susceptibles de voir leurs publications comparé à **Nike** qui a le plus d'*abonnés* (8.208.439).

Malgré le peu de publications postées par **Adidas** (voir section 3.1.1), la marque a quand même pu acquérir un nombre conséquent d'*abonnés* (3.788.472). Ceci peut être dû au fait que la marque possède déjà une certaine notoriété ou parce que ses publications sont intéressantes même si peu nombreuses. En effet, on peut aussi constater que son nombre moyen de *favorites* est élevé (790). C'est cependant le contraire pour **Salomon** qui a un grand nombre de publications mais le plus petit nombre d'*abonnés*.

⁴ Le nombre d'abonnés (followers) est le nombre de personnes qui s'abonnent à la page pour pouvoir voir leurs publications dans leur propre fil d'actualité

⁵ Le nombre de followers pour Asics peut être légèrement biaisé car il n'y a pas de page générale mais simplement des pages par région. Ici, la région est dénommée « America ».

3.2. Analyse de sentiments

Après avoir fait une première analyse descriptive des données, une analyse des sentiments présents dans les textes a été menée en suivant les recommandations faites dans la littérature. Les sentiments véhiculés par les pages des marques seront ainsi dévoilés.

3.2.1. Principes de la reconnaissance des émotions

La reconnaissance des émotions repose sur un algorithme de machine learning devant être entraîné avec des données déjà classifiées pour pouvoir ainsi prédire la classification de nouvelles données. En effet, le modèle doit tout d'abord apprendre à classifier les données et ce en analysant à quel type de phrase est assigné quel type de classe. Dans notre cas, les classes seront les différents sentiments. Ensuite, le modèle peut ainsi classer lui-même des données auxquelles aucun sentiment n'a encore été attribué. Ici, l'algorithme choisi est « Naive Bayes Classifier ». L'algorithme du Naive Bayes Classifier est présent dans au moins deux bibliothèques Python d'analyse de données. La première est TextBlob qui est une librairie de traitement de données textuelles. L'algorithme de cette librairie ne sera pas utilisé car son temps de traitement n'est pas adapté à la taille de nos datasets. De plus, l'entraînement de ce modèle ne peut pas se faire avec un grand nombre de données classifiées. Hors, au plus le modèle reçoit de données pour son entraînement, au mieux il est entraîné pour prédire de nouvelles données. Le modèle présent dans la librairie Scikit-Learn a donc été choisi et plus précisément le « Multinomial Naive Bayes » qui est un classificateur qui convient à la classification pour des données distribuées multinomialement. Ce sont des données ayant plus de deux classes. En effet, nous allons voir juste après que nos trois datasets d'entraînement contiennent entre sept et treize classes différentes. Les avantages de ce modèle sont que son traitement est rapide et qu'il peut s'entraîner sur une large base de données.

Pour notre analyse, trois bases de données d'entraînement différentes seront utilisées pour pouvoir se baser sur trois résultats différents et ainsi confirmer ou infirmer ce que chacune produit. De plus, étant donné que les résultats fournis peuvent paraître déroutants, il est important de comprendre si ceci est dû à un problème de donnée d'entraînement ou si cela découle simplement des données collectées sur Twitter. Finalement, il faut savoir qu'un tel type de base de données n'est pas facile à trouver. Ce ne sont donc pas toujours les bases de données souhaitées qui sont trouvées. Plusieurs analyses ont donc été faites pour vérifier si les résultats produits par chaque analyse sont cohérents.

Le premier set d'entraînement a été trouvé sur le site data.world⁶ mais provient de la plateforme créée pour les équipes de data science, Crowdfunder. Le dataset compte 39 811 lignes, chaque ligne comprenant un tweet et son sentiment assigné. Treize types de sentiments différents sont utilisés pour classer ces données, à savoir la colère, l'ennui, le vide, l'enthousiasme, l'amusement, le bonheur, la haine, l'amour, la neutralité, le soulagement, la tristesse, la surprise et l'inquiétude. La FIGURE 06 représente la proportion de chaque émotion dans le fichier. Comme on peut le voir, certains sentiments comme l'*inquiétude* ou la *neutralité* sont plus représentés que d'autres dans cet ensemble d'entraînement. Le *bonheur* et la *tristesse* sont aussi présents en plus grande quantité.

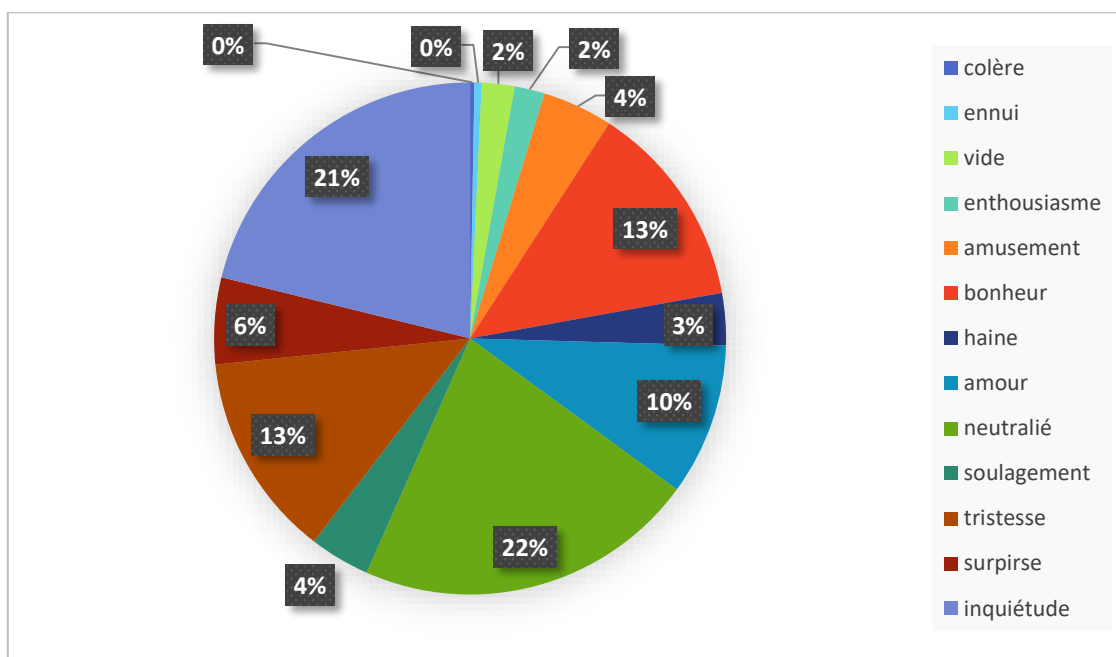


FIGURE 06 – Graphique à secteur représentant le pourcentage d'émotions dans le 1^{er} dataset

Le deuxième ensemble d'entraînement [21] comprend quant à lui sept émotions, la colère, la peur, la joie, l'amour, la neutralité, la surprise et la tristesse. Cet ensemble est moins fourni que le premier car il ne comporte que 4791 données. De plus, ces données ne sont pas des tweets labélisés avec une émotion mais ce sont des textes venant de commentaires du site d'entraide pour la programmation, « Stack Overflow ». Dans cet ensemble, certaines émotions sont aussi surreprésentées, il s'agit de l'amour et de la neutralité qui représentent ensemble environ 66% de tous les éléments.

La dernière base de données [22] comprend aussi sept classes différentes mais deux d'entre elles ne sont pas vraiment pertinentes, ce sont les classes, « pas de code » et « pas pertinent ». Les autres classes sont la colère, le dégoût, le bonheur, la tristesse et la surprise. On peut remarquer que certaines émotions reviennent dans tous les ensembles d'entraînement comme le bonheur ou la joie, la colère, la surprise, la tristesse, la neutralité. Excepté la neutralité qui n'est pas une émotion en soi, ce sont les

⁶ https://data.world/crowdfunder/sentiment-analysis-in-text/workspace/file?filename=text_emotion.csv

émotions de base selon Ekman (1992) [23]. Ce dataset est encore plus petit que le précédent, avec 2782 tweets associés à des émotions. Les émotions surpondérées sont ici le bonheur et « pas de code » qui représentent ensemble presque 90% du fichier, ce qui n'est pas négligeable.

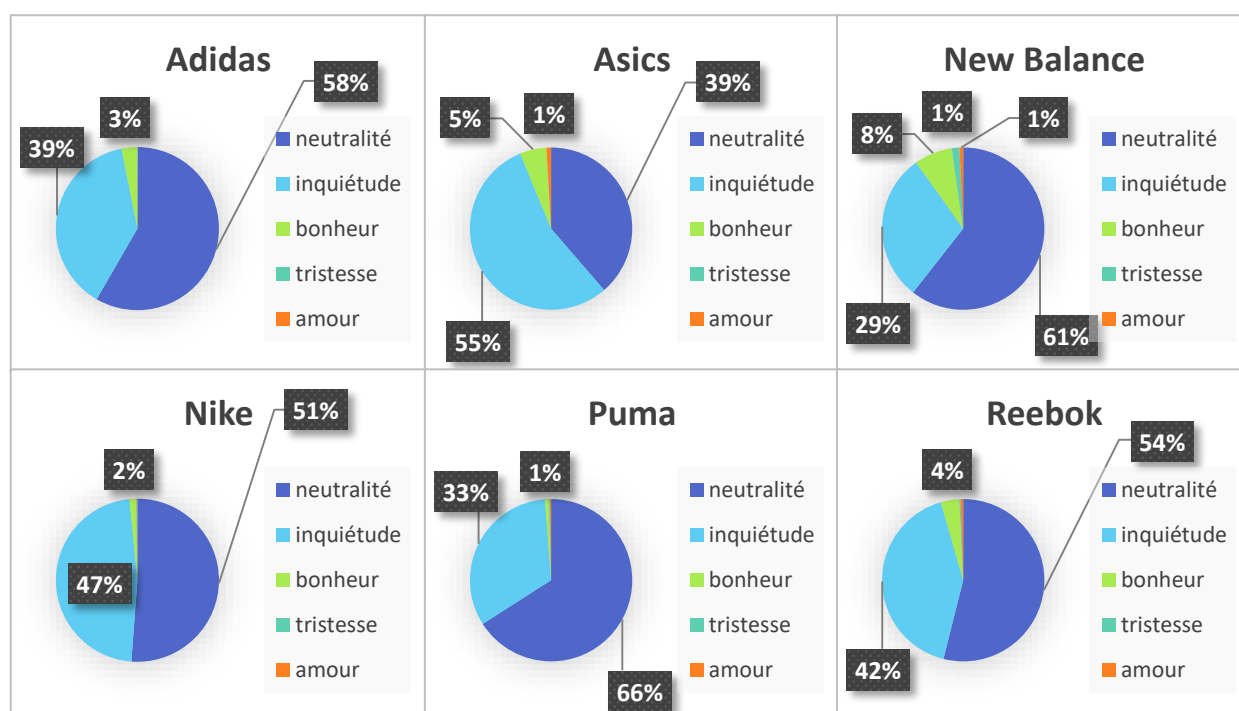
L'algorithme de Naives Bayes Classifier va donc être entraîné une première fois puis donnera les résultats des sentiments de notre base de données « *twitter_pages* ». Il sera ensuite réinitialisé et entraîné une deuxième fois pour le deuxième dataset, puis les résultats apparaitront et la procédure sera la même pour le troisième ensemble d'entraînement.

3.2.2. Résultats

Le TABLEAU 03 représente les résultats de l'analyse de sentiments avec le premier set d'entraînement. Il affiche le nombre de tweets qui tombent dans une certaine catégorie d'émotions pour les différentes marques. La FIGURE 07 représente les mêmes résultats mais sous forme de pourcentage.

Sentiment	Adidas	Asics	New Balance	Nike	Puma	Reebok	Salomon	The North Face	Under Armour
neutral	56	256	80	560	1870	206	1093	235	218
worry	37	365	39	517	931	159	705	118	200
happiness	3	35	10	14	22	14	251	7	12
sadness	0	0	2	2	5	1	2	1	0
love	0	6	1	1	7	2	22	9	0

TABLEAU 03 – Nombre d'émotions par marque avec le 1^{er} ensemble de données



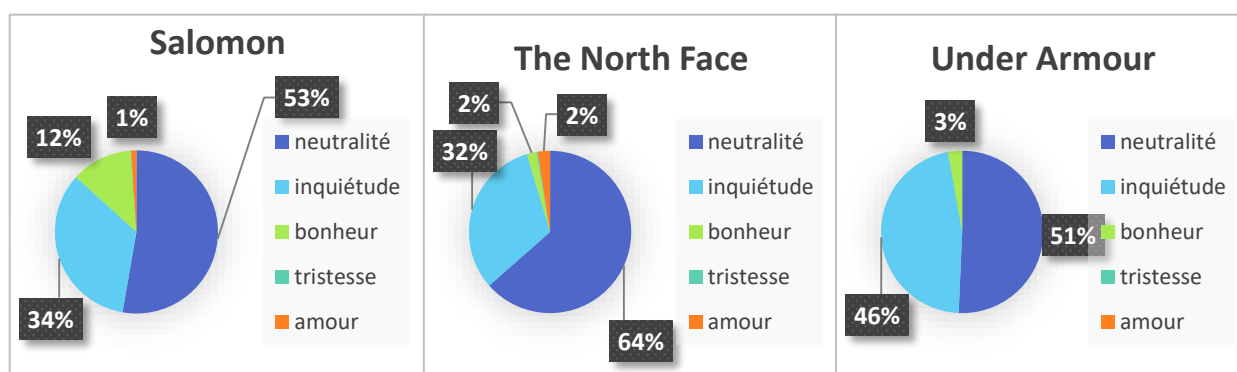


FIGURE 07 – Graphiques représentant la proportion d'émotions par marque avec le 1^{er} ensemble de données

Dans le TABLEAU 03 et la FIGURE 07, on peut observer que parmi toutes les marques, seulement cinq sentiments sur les treize de départ ont été identifiés, à savoir la **neutralité**, l'**inquiétude**, le **bonheur**, la **tristesse** et l'**amour**. Il s'agit des sentiments les plus présents dans l'ensemble d'entraînement.

Au vu de la surreprésentation des catégories **inquiétude** et **neutre** dans le dataset d'entraînement, le fait que ces deux sentiments ressortent dans les résultats est assez logique. Ce qui peut être intéressant, c'est de comparer les marques entre elles par rapport à ces sentiments mais aussi par rapport aux deux ou trois autres sentiments qui apparaissent également. On peut noter qu'**Asics** a plus de sentiments d'**inquiétude** (55%) que les autres tandis que c'est **New Balance** (29%) qui en a le moins. Les autres marques se situent toutes dans les alentours de 35%, ce qui est tout de même plus élevé que le pourcentage de 22% dans l'ensemble d'entraînement (voir FIGURE 06).

La **neutralité** ne sera pas étudiée ici mais il peut être intéressant de s'attarder légèrement sur le **bonheur**. **Salomon** (12%) est la page qui produit le plus de tweets qui y font référence, suivi par **New Balance** (8%) et finalement **Asics** (5%). Avec seulement 1%, c'est **Puma** qui utilise le moins ce sentiment dans ses tweets.

Finalement, le sentiment d'**amour** est légèrement présent chez **The North Face** (2%) mais aussi chez **New Balance**, **Salomon** et **Asics** avec seulement 1% pour ces quatre dernières marques.

Les résultats produits avec les deux autres ensembles de données sont assez similaires car il n'y a que deux sentiments qui se démarquent, la **neutralité** et l'**amour** pour le deuxième et le « **pas de code** » et le **bonheur** pour le troisième. Cette constatation nous conforte dans l'idée que les sets d'entraînement produisent tout de même de bons résultats. La FIGURE 08 présente les résultats pour le deuxième ensemble de données et la FIGURE 09 pour le troisième.

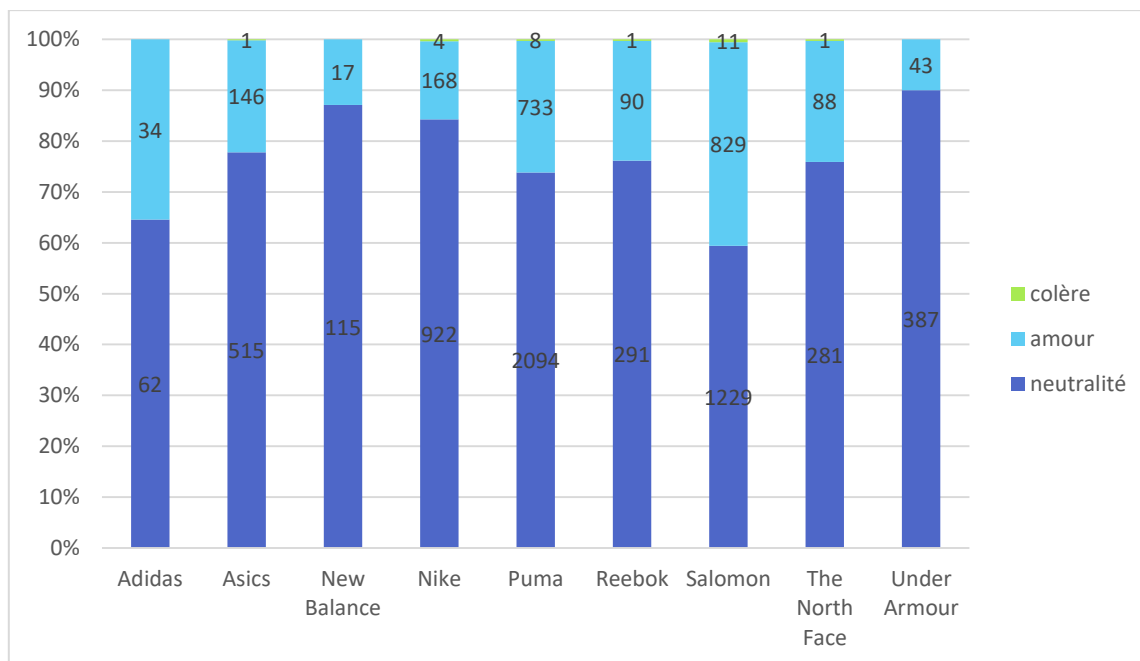


FIGURE 08 - Graphique représentant la proportion d'émotions par marque avec le 2^{ème} ensemble de données

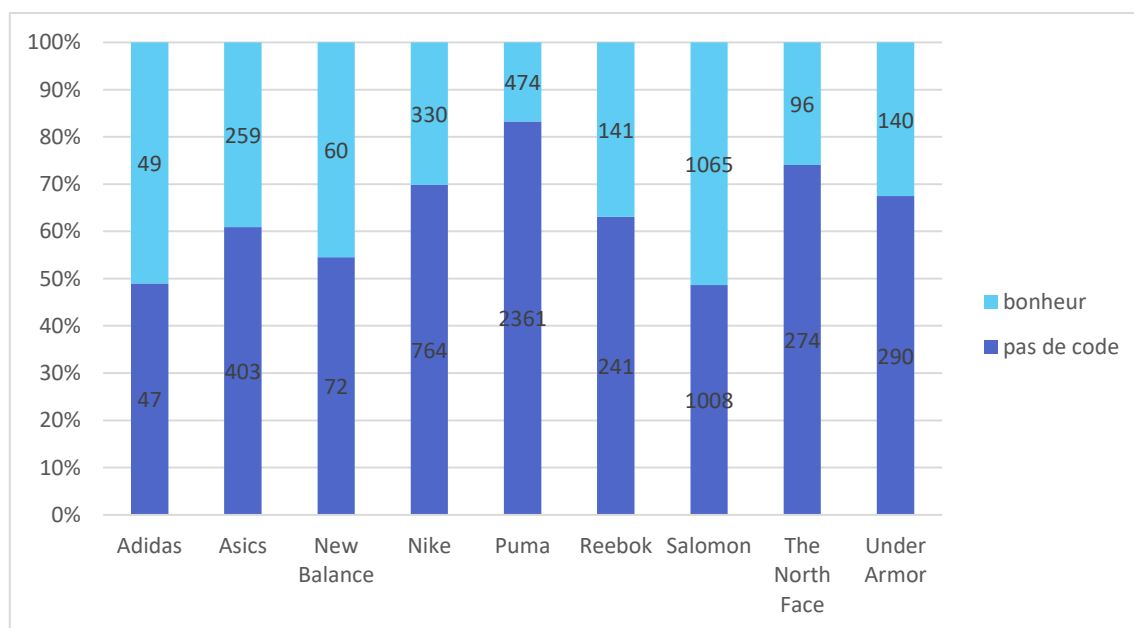


FIGURE 09 - Graphique représentant la proportion d'émotions par marque avec le 3^{ème} ensemble de données

Dans ces graphiques, on peut voir que c'est la portion « non classifiée » ou neutre qui a la part la plus importante, allant jusqu'à environ 90% pour **Under Armour** avec le deuxième ensemble d'entraînement. Il n'y a que pour **Adidas** et **Salomon** pour lesquels la proportion chute en-dessous de 50% avec dans la FIGURE 09.

Ici, on fait l'hypothèse que les sentiments d'**amour** et de **bonheur** sont assez similaires et qu'ils ont une dénotation positive. On peut donc comparer les deux résultats en émettant l'hypothèse que ce

sont les mêmes sentiments. Les résultats apparaissent en proportion presque similaire excepté que les pourcentages du sentiment de **bonheur** sont un peu plus élevés que pour l'**amour**, excepté pour **The North Face** ou les proportions sont assez identiques, pour **Puma** dont l'**amour** est plus élevé et pour **New Balance** pour qui le **bonheur** est bien plus présent. Sinon, la tendance avec **Salmon** ayant le plus de sentiments positifs suivi par **Adidas** et **Asics** est présente dans les deux résultats. **Nike**, **Reebok** et **Under Armour** étant un peu derrière et plus ou moins tous les trois au même niveau.

Dans le résultat avec le deuxième ensemble de données, on peut aussi voir apparaître légèrement le sentiment de **colère** pour **Nike**, **Puma** et **Salomon** qui sont des pages postant beaucoup de tweets et c'est donc pour ça que ce sentiment apparaît.

Ces trois méthodes présentent des résultats relativement différents en ce qui concerne le sentiment d'inquiétude pour la première étude et le sentiment positif pour les deuxième et troisième. Il n'est donc pas intéressant de s'attarder précisément sur les chiffres qu'elles produisent. Cependant, il sera intéressant de comparer ces résultats avec ceux des études identiques pour les données relatives à la communauté des marques.

3.3. Analyse du moment des publications

Les moments privilégiés par les marques pour publier des tweets peuvent aussi nous apporter des informations quant à leurs différentes stratégies. Dans la section 3.1.1, nous avons déjà noté que des marques telles que **Puma**, **Salomon** et **Nike** postent plus de publications, comparé à Adidas et **New Balance** qui en postent le moins. Ici, nous allons plutôt analyser à quel moment ces marques postent le plus souvent, que ce soit en fonction de la semaine pendant l'année, du jour pendant la semaine ou même de l'heure pendant la journée mais aussi en fonction des grands événements sportifs. En effet, selon Elizabeth Arens (2020) [24], pour qu'une publication puisse s'élever au-dessus du bruit des réseaux sociaux et recevoir de l'attention, il y a des conseils à suivre. Le contenu est un aspect important pour pouvoir se démarquer mais le moment auquel la publication est postée joue aussi un rôle. Dans son article, elle identifie le moment optimal pour chaque genre de contenu mais aussi pour chaque réseau social. L'ANNEXE 02 présente ses résultats pour les biens de consommation sur Twitter. Ce graphique révèle que le moment idéal pour publier un post est le dimanche à 11h. En termes de jours, c'est le dimanche le meilleur et le lundi le pire. En termes d'heures, les plus favorables se trouvent entre 9h et 16h.

3.3.1. Semaine et événements marquants

Les FIGURES 10, 11 et 12 représentent les résultats de l'analyse des dates de publications en fonction de la semaine durant l'année. « 1 » représentant la première semaine, c'est-à-dire du 31 décembre 2018 au 6 janvier 2019 et ainsi de suite.

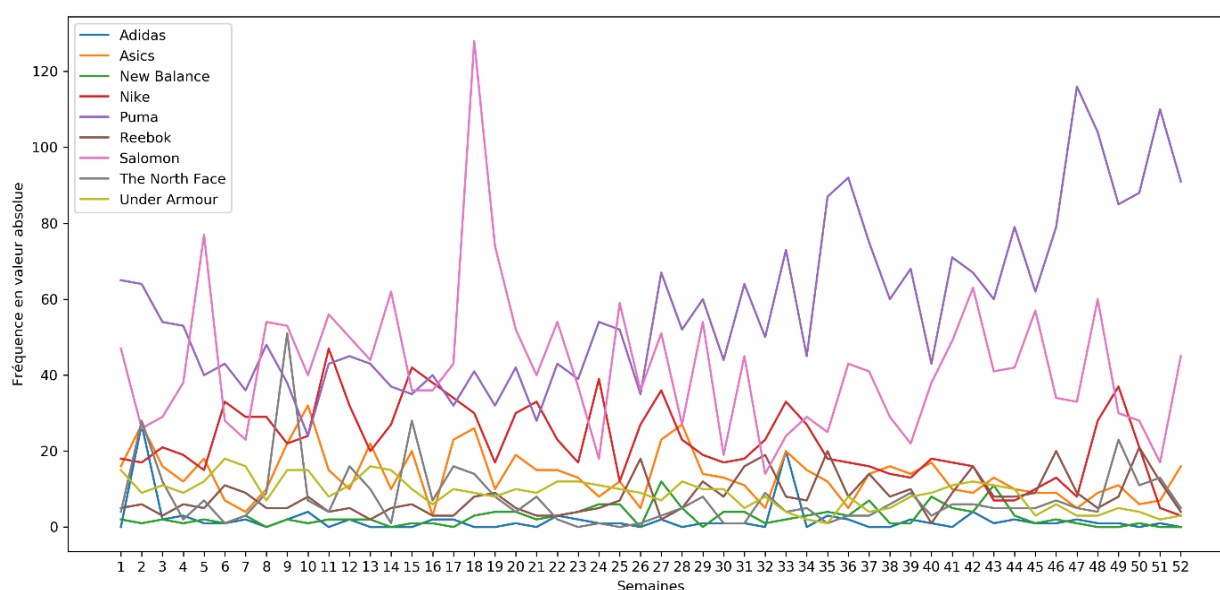


FIGURE 10 – Graphique représentant le nombre de tweets en valeur absolue par marque et en fonction de la semaine

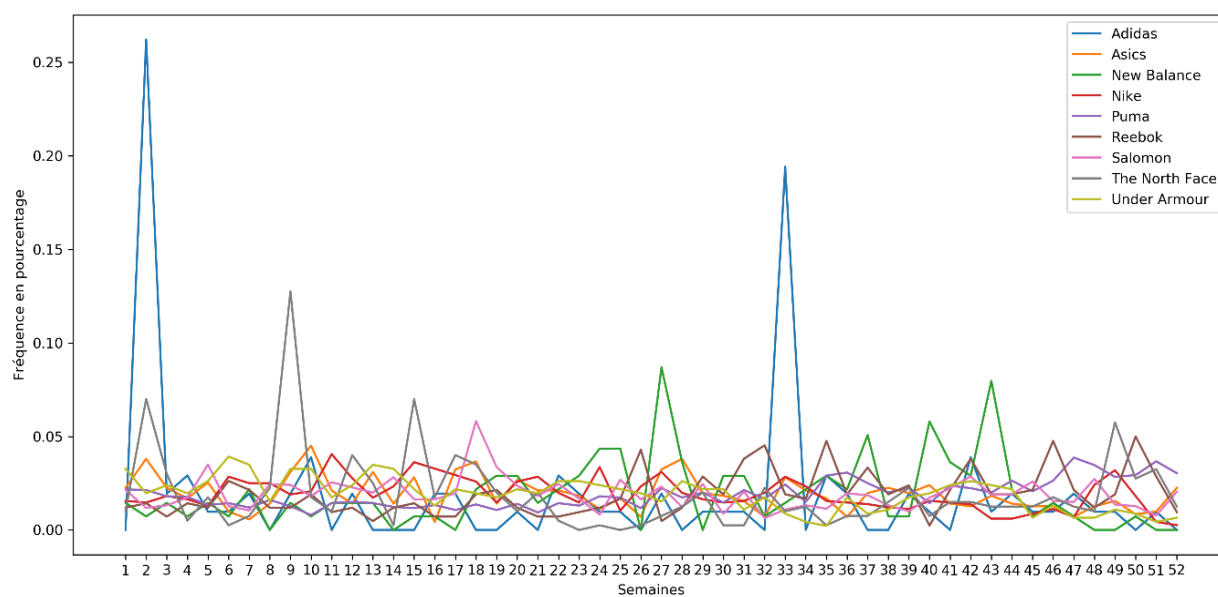


FIGURE 11 – Graphique représentant le nombre de tweets en pourcentage par marque et en fonction de la semaine

Dans le graphique suivant, **Adidas** a été retiré pour pouvoir observer les autres marques avec un niveau de granularité plus approprié.

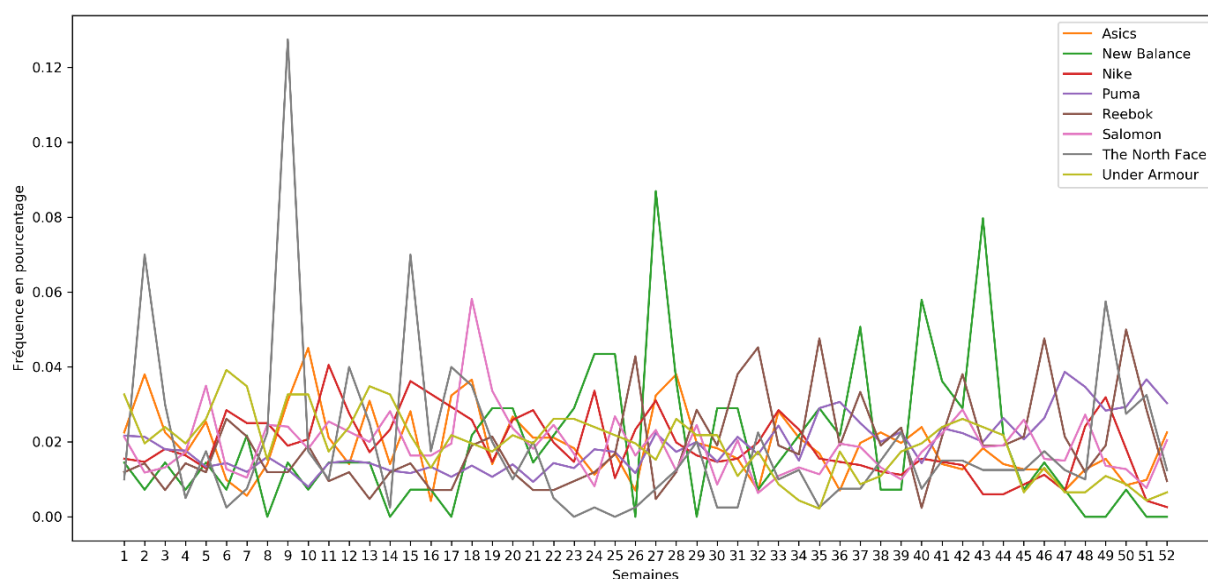


FIGURE 12 – Graphique représentant le nombre de tweets en pourcentage (excepté Adidas) et en fonction de la semaine

Dans ces graphiques, **Adidas**, **The North Face**, **Reebok** et **New Balance** ressortent car ils fonctionnent sous forme de « pic », il y a certaines semaines lors desquelles ils postent bien plus en comparaison aux autres semaines. Les autres marques sont plus constantes au cours de l'année. **Salomon** possède cependant un pic lors de la 18^{ème} semaine qui correspond à la semaine du 29 avril au 5 mai. En regardant de plus près aux tweets de cette semaine, il s'agit simplement d'un taux de réponses plus élevé de la part de Salomon à ses followers. Ces réponses sont des précisions pour le

concours « SalomonWMN » (voir section 3.1.2), des encouragements aux sportifs, des remerciements pour le support ou encore des réponses de service-après-vente. Le plus grand pic pour **The North Face** correspond à la date à laquelle le film « Free Solo » a gagné un Oscar. Ce film parle de la quête du sommet El Captain sans corde par un célèbre grimpeur, Alex Honnold. Alex est un athlète sponsorisé par The North Face. Les premier et troisième pics sont aussi une augmentation du taux de réponses. Ici, les réponses sont majoritairement pour le service-après-vente. La marque s'excuse pour des problèmes subis par les utilisateurs, d'où la fréquence du mot « *sorry* » dans le TABLEAU 01. Ensuite, elle fait des recommandations aux utilisateurs pour qu'on les aide à régler leur problème. Le pic le plus élevé de **New Balance** correspond à un chevauchement du tournoi de tennis de Wimbledon et à la coupe du monde féminine de football. Le premier pic d'**Adidas** est lié au hashtag #CreatorsUnite qui est un mouvement réunissant des créateurs pour créer l'inattendu. Le deuxième pic d'Adidas correspond au 18 août qui est la date de création d'Adidas et pour laquelle la marque a proposé un petit historique en plusieurs tweets à ses followers.

Nous avons pu constater dans les graphiques précédents que les pics de tweets ne correspondaient pas vraiment à de grands événements sportifs mais analysons tout de même si ces événements impactent la communication des marques. Pour ce faire, nous allons confronter tous les événements listés en ANNEXE 03 avec les mots présents dans les tweets et leur fréquence.

Parmi les grands événements sportifs, ils sont peu nombreux à impacter la communication des marques. Le plus important est le *marathon de Boston* du 15 avril 2019. C'est le seul à impacter deux marques différentes, à savoir **Asics** et **Salomon**. Cependant, l'impact n'est pas très important car on ne retrouve par marque qu'environ cinq tweets parlant de cet événement. Les autres événements ayant un impact mais tout aussi léger sont les *championnats d'athlétisme aux Etats-Unis* et le *tournoi de tennis de Wimbledon* pour **New Balance**, le *Tahoe Rim Trail*, le *Pikes Peak Marathon* et le *SalomonWMN camp* pour **Salomon**. Le dernier événement fait partie d'une campagne qui faisait gagner à dix femmes un camp au Monténégro accompagnées par des personnalités féminines liées à Salomon. La marque a aussi suivi la saison du *Golden Trail World Series* qui est une compétition de sept grandes courses sponsorisée par la marque. Comme vu dans les mots les plus fréquents de **Reebok**, les *CrossFit Games* sont présents dans la communication de la marque. Finalement, **Under Armour** a suivi la saison de la ligue majeure de baseball en Amérique du nord.

Tous ces événements ont été relevés dans le contenu des tweets des marques mais n'ont pas un fort impact. En effet, entre cinq et treize tweets ont été relevés pour ces événements. De plus, les douze tweets sont pour la ligue de baseball qui se déroule pendant plus de six mois. C'est donc pour cela qu'aucun pic n'apparaît dans les FIGURES 10, 11 ou 12 pour ces événements.

3.3.2. Jour de la semaine

Les FIGURES 13 et 14 représentent les résultats de l'analyse du moment des publications en fonction des jours de la semaine. Le trait en pointillé noir représente le nombre moyen de mention « favorite » qu'ont reçu toutes les publications des marques.

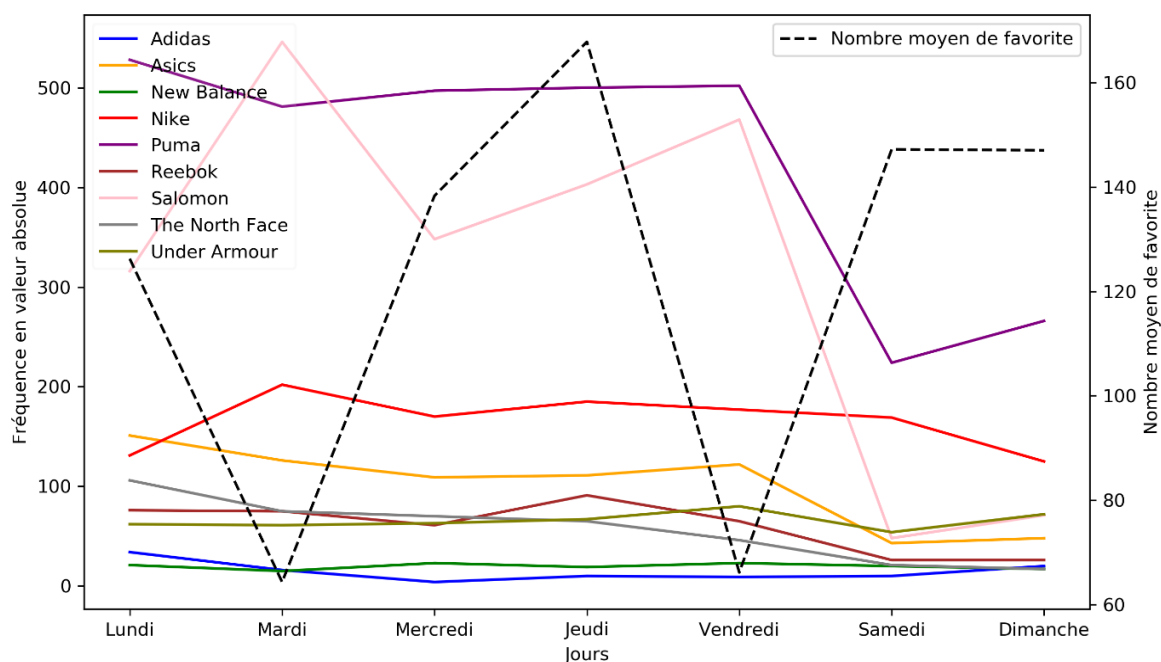


FIGURE 13 – Graphique représentant le nombre de tweets en valeur absolue par marque et le nombre moyen de « favorite » en fonction du jour de la semaine

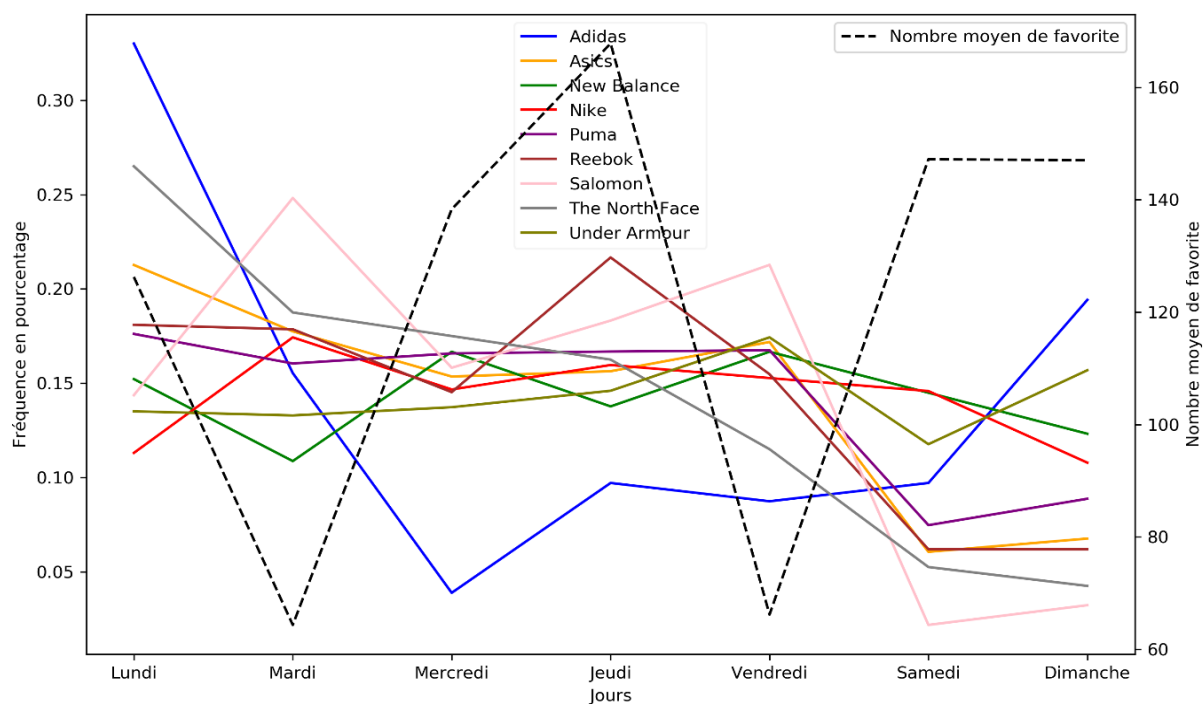


FIGURE 14 – Graphique représentant le nombre de tweets en pourcentage par marque et le nombre moyen de « favorite » en fonction du jour de la semaine

Ce qui ressort le plus nettement des deux graphiques précédents est qu'en général, les marques postent moins le *samedi*, excepté peut-être **Nike** et **New Balance**. Ceci est probablement lié au fait que les employés ne travaillent pas le samedi et par conséquent, ne savent pas poster sur les réseaux sociaux. Pour **Salomon** et **Puma**, cette différence est très visible dans le premier graphique. Cependant, les posts sont plus nombreux le *dimanche* que le samedi alors que les employés sont aussi supposés être en week-end et donc ne pas pouvoir poster sur les réseaux. Les marques ont peut-être remarqué que les tweets postés le dimanche obtiennent en moyenne beaucoup de « J'aime » et ce serait pour cela qu'ils engagent quelques employés pour poster des tweets le dimanche. La plupart des utilisateurs de Twitter sont aussi en congé pendant le week-end et ont donc plus de temps à passer sur les réseaux sociaux. Ils peuvent par conséquent s'attarder sur les tweets des marques et y apposer une mention « J'aime ». Il pourrait être opportun de faire de même avec le samedi car c'est aussi une journée lors de laquelle les « J'aime » sont nombreux.

Ensuite, on peut affirmer que le nombre de tweets pour les jours de la semaine est assez constant hormis pour **Adidas** qui a choisi, volontairement ou non de privilégier le *lundi* pour ses posts et de mettre le *mercredi* de côté. Cependant, Adidas n'étant pas une marque postant beaucoup de contenus sur Twitter, cet effet est peut-être simplement dû au hasard. D'autres marques comme **Asics**, **Reebok**, **The North Face** ou **Puma** privilégient aussi le *lundi* pour publier des tweets, le nombre de « J'aime » n'étant pourtant pas supérieur à la moyenne en ce jour.

Pour nos marques de sport, le *jeudi* est le jour où leurs publications obtiennent le plus de mention « favorite » et le *mercredi* est un bon jour aussi. Ces constatations sont en raccord avec l'ANNEXE 02 où le *dimanche* et le *mercredi* sont les jours où les tweets reçoivent le plus d'attention. Il est donc opportun pour les entreprises de changer leur moment de publications sur Twitter pour privilégier ces jours.

3.3.3. Heure

Les FIGURES 15 et 16 représentent les résultats de l'analyse de l'instant des posts en fonction de l'heure pendant la journée. Ici, l'heure est UTC. Les pages des marques étant internationales, l'heure en elle-même n'est pas très significative car il n'est pas la même heure partout lorsqu'une page poste une publication. Cependant, nous pouvons tout de même analyser les tendances et comparer le nombre de publications par heure avec le nombre de mention « favorite ». De plus, nous pourrions constater que la fréquence en fonction des heures peut faire penser que l'heure en UTC correspond au fuseau horaire de l'heure des posts car on peut y retrouver le déroulement d'une journée normale.

Afin de comparer le nombre de « favorite » en fonction de l'heure, il faut savoir que les mentions peuvent être apposées à n'importe quel moment après qu'un post ait été publié. Cependant, selon

Spasojevic, Li, et al. (2015) [25], une majorité des réactions se produit dans les deux heures après le moment de publication. En effet, au plus le post est récent et au plus il a de chance d'apparaître dans le fil d'actualité des utilisateurs. De plus, Twitter est un réseau social qui a un temps de réaction plus court qu'un autre comme Facebook par exemple.

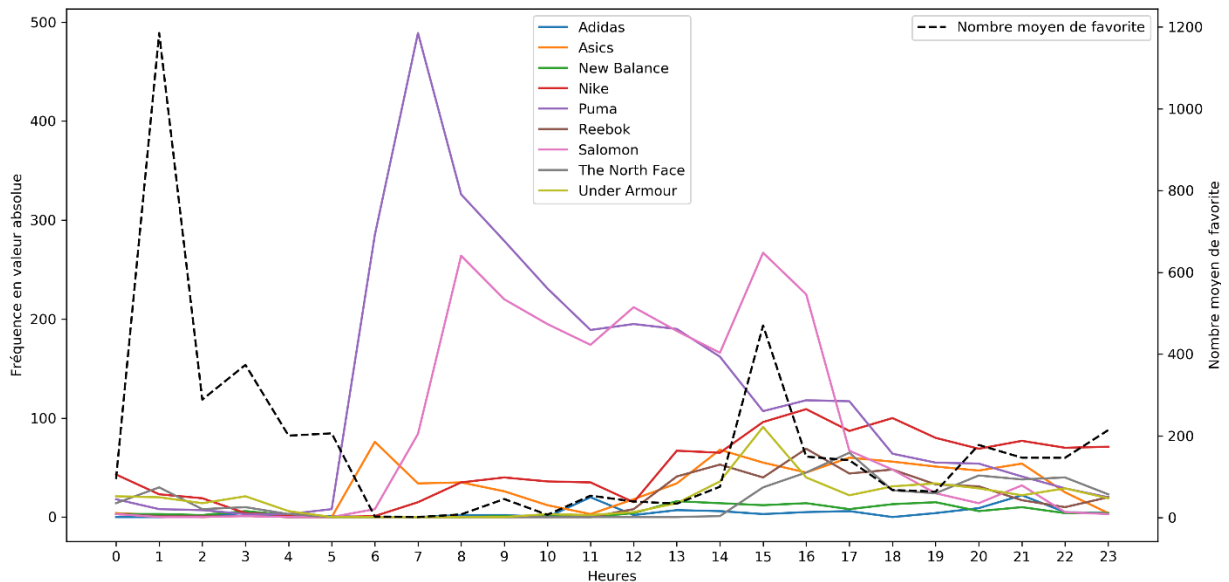


FIGURE 15 – Graphique représentant le nombre de tweets en valeur absolue par marque et le nombre moyen de « J'aime » en fonction de l'heure

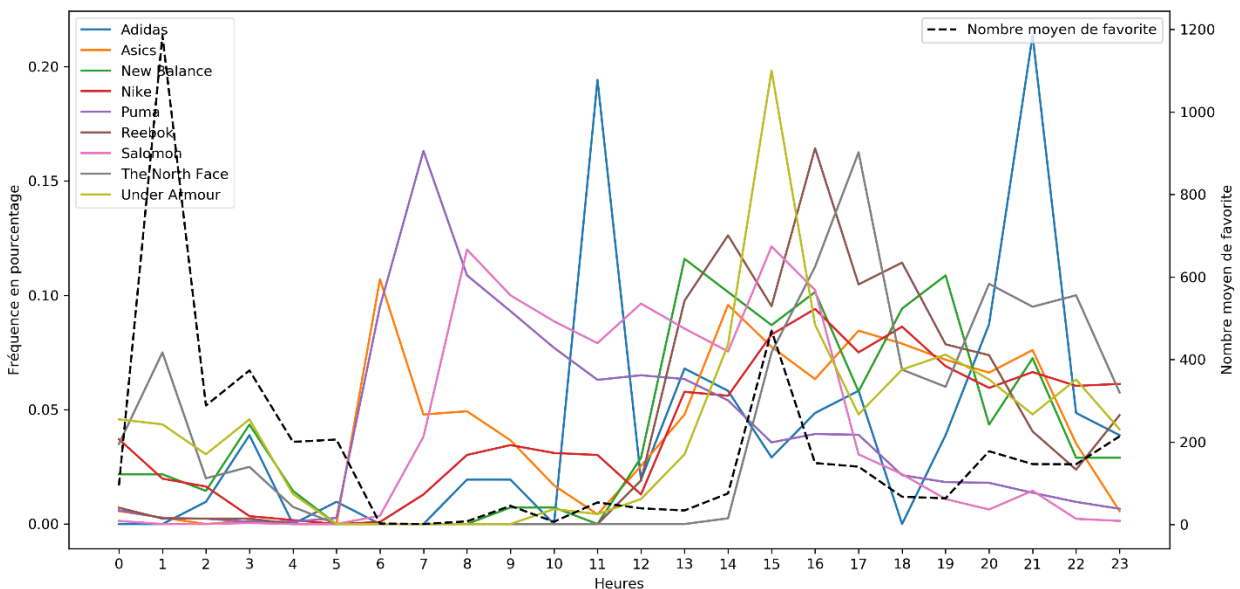


FIGURE 16 – Graphique représentant le nombre de tweets en pourcentage par marque et le nombre moyen de « J'aime » en fonction de l'heure

Sur ces graphiques, on peut voir que les tweets ne sont pas souvent postés avant 5h du matin UTC, ce qui est assez sensé car il y a beaucoup moins d'utilisateurs présents sur les réseaux sociaux pendant la nuit. De plus, les employés des différentes marques ne sont pas souvent disponibles pour créer des tweets pendant la nuit. On peut assister à un décollage du nombre de tweets dans les

environs de midi et une plus forte fréquence vers 15-16h UTC pour commencer à diminuer après 22h UTC. Toutes ces fluctuations sont assez sensées au vu du déroulement d'une journée type.

En ce qui concerne le nombre de « favorite », on observe deux pics. Le premier se situant pendant la nuit, entre 1h et 5h UTC, ne va pas être considéré ici. En effet, le nombre de tweets est trop faible à ce moment pour définir une tendance. Le pic à 1h du matin UTC est majoritairement dû à un seul tweet posté par **Puma** avec le texte « Rest In Peace » en hommage à Nipsey Hussle, un rappeur qui avait un partenariat avec Puma et qui a été assassiné quelques heures avant la publication. Ce tweet à lui seul a récolté 66,169 « favorite ». Un seul tweet ne peut donc pas définir une tendance.

Le deuxième pic du nombre de « favorite » qui se situe entre 14h et 17h UTC, se range dans les mêmes heures que le pic de fréquence de tweets ce qui laissent penser que les marques ont compris à quelles heures il était plus favorables pour elles de poster une publication. Il y a aussi une légère augmentation entre 20h et 23h UTC même si le nombre de tweets a déjà diminué à ces heures.

Même si l'heure est prise en UTC, la fréquence des tweets est assez concordante avec le déroulement d'une journée normale. De plus, comme nous le verrons dans la section 5.3, les heures les plus favorables selon l'étude d'Elizabeth Arens (2020) [24] confirme notre hypothèse de se concentrer seulement sur le deuxième pic de mention « favorite ».

3.4. Analyse de stratégie de communication

Pour pouvoir analyser les différentes stratégies de communication telles que définies par Coursaris et Osch [15] (voir section 2.1), il a fallu trouver une méthode pour identifier à quelle stratégie appartient un certain tweet. De plus, contrairement à Coursaris et Osch qui ont passé chacune des publications en revue, notre méthode devait être automatisée car ce n'est pas efficace de lire chacun des 8000 tweets. Pour cela, des indicateurs de substitutions (*proxy indicators*) ont été utilisés. Il s'agit d'indicateurs qui servent d'estimateurs pour la variable selon laquelle un tweet se rapporte à une stratégie. Ici, pour chacune des stratégies des listes de mots considérés comme révélateurs de la stratégie ont été établies (voir ANNEXE 04) et constituent donc ces indicateurs de substitution. Ces listes comprennent des mots qui peuvent nous rapprocher d'un certain type de communication. Elles ont été construites de la façon suivante :

1/ Tout d'abord, dans leur article, Coursaris et Osch [15] présentent quelques exemples de publications pour chaque type de stratégie. Les mots importants de ces exemples ont donc été mis dans les listes.

2/ Ensuite, tous les mots présents dans les tweets des pages et ayant une fréquence de plus de cinq occurrences ont été examinés. Les mots qui se rapportaient à une certaine stratégie ont été mis dans la liste relative à celle-ci.

3/ Pour la communication en rapport avec la **notoriété des produits**, la liste a été complétée par les noms de tous les produits vendus par la marque. Les mots de cette catégorie ont été choisis grâce à une évaluation des sites internet de la marque dans lesquels ils mettent tous les articles en vente. Le nom des catégories de chaque produit a ainsi été relevé. Par exemple, le mot « air » a été ajouté à la liste des produits de **Nike** en raison de la célèbre catégorie de baskets Nike Air.

4/ Finalement, les événements cités dans la section 3.3.1 ont été ajoutés à la liste de la stratégie **saisonnière** car celle-ci inclut les posts en rapport avec des événements annuels.

5/ L'indicateur de substitution pour les stratégies **service client** et **engagement** ne se limite pas seulement à la liste de mots présente en ANNEXE 04. En effet, il y a une grande partie des tweets récoltés sur les pages qui sont des réponses (*replies*) aux utilisateurs. Ces tweets sont donc considérés comme relevant du service client ou de l'engagement. Notre méthode va donc repérer les tweets de réponses et les affecter à l'une de ces stratégies.

Si un mot d'une certaine liste est présent dans un tweet alors la stratégie correspondant à la liste sera celle désignée pour ce tweet. Pour que la méthode soit pertinente, les listes de mots sont testées dans un certain ordre :

1. Responsabilité sociale des entreprises
2. Nom de célébrité pour la notoriété de la marque

3. Notoriété des produits
4. Promotion
5. Saisonnier
6. Service client + engagement
7. Nom de marque pour la notoriété de la marque

Il ne serait en effet pas logique de comparer un tweet avec la liste comprenant le nom des marques en premier car celui-ci est susceptible de ressortir pour d'autres stratégies, il faut donc tester cette liste en dernier lieu et seulement si aucune autre catégorie n'a été attribuée avant.

Comme vous pouvez le voir dans l'ordre pour tester les stratégies, le **service client** et l'**engagement** ont été rassemblés. En effet, il n'était pas possible de les dissocier. Ces deux catégories se rapportent à des réponses que peut apporter la marque à ses consommateurs. En ce qui concerne le service client, ce sont plutôt des réponses techniques ou des réponses à des plaintes et en ce qui concerne l'engagement, c'est plutôt des réponses qui montrent que la marque s'intéresse à ce que ses consommateurs. Nous avons pu différencier les tweets de réponses des tweets « normaux » mais la distinction dans le type de réponses n'a pas pu être faite et c'est pour cette raison que les deux catégories ont été rassemblées. Cependant, ce n'est pas un problème car ce sont tout de même des catégories assez similaires.

Les TABLEAUX 04 et 05 présentent les résultats de cette analyse en donnant en nombre absolu ou en pourcentage la proportion des tweets par stratégie et par marque.

	Adidas	Asics	New Balance	Nike	Puma	Reebok	Salomon	The North Face	Under Armour
Notoriété de la marque	7	26	42	4	162	45	365	11	127
Responsabilité sociale	8	7	0	10	5	7	52	42	11
Service client + engagement	41	280	35	815	2361	162	1243	151	57
Notoriété des produits	12	177	10	192	144	82	161	54	81
Promotion	2	44	3	35	5	6	100	8	7
Saison	1	12	11	9	8	7	57	10	25
Aucun	25	116	31	29	150	73	95	94	122
TOTAL	96	662	132	1094	2835	382	2073	370	430

TABLEAU 04 – Nombre de tweets en valeur absolue par catégorie et par marque

	Adidas	Asics	New Balance	Nike	Puma	Reebok	Salomon	The North Face	Under Armour	TOTAL
Notoriété de la marque	7,29	3,93	31,82	0,37	5,71	11,78	17,61	2,97	29,53	12,33
Responsabilité sociale	8,33	1,06	0,00	0,91	0,18	1,83	2,51	11,35	2,56	3,19
Service client + engagement	44,79	44,56	28,03	78,61	87,58	44,76	63,43	42,43	14,19	49,82
Notoriété des produits	12,50	26,74	7,58	17,55	5,08	21,47	7,77	14,59	18,84	14,68
Promotion	2,08	6,65	2,27	3,20	0,18	1,57	4,82	2,16	1,63	2,73
Saison	1,04	1,81	8,33	0,82	0,28	1,83	2,75	2,70	5,81	2,82
Aucun	26,04	17,52	23,48	2,65	5,29	19,11	4,58	25,41	28,37	16,94

TABLEAU 05 – Pourcentage de tweets par catégorie et par marque

Grâce aux informations que contiennent ces tableaux, on peut commencer à comprendre comment se comportent les différentes marques de sport sélectionnées en fonction de leur stratégie de communication.

L'ensemble des tweets relève majoritairement de l'association **service client + engagement** (49.82%), spécialement pour **Nike** et **Puma** pour lesquels ces taux atteignent 78.61% et 87.58% respectivement. Dans le premier tableau, on peut d'ailleurs apercevoir que ces deux marques ont un des nombres de tweets les plus élevés et que cela est majoritairement dû à ces tweets. Ces tweets sont en réalité des réponses à certains utilisateurs. Pour **Under Armour**, cette catégorie est nettement plus faible que les autres avec seulement 14.19%, ce qui signifie que Under Armour ne prend pas le temps de répondre à ses followers sur Twitter. En effet, comme on peut le voir sur la FIGURE 05, Under Armour n'a émis que cinq réponses parmi tous ses tweets. La marque préfère se mettre en avant avec des tweets augmentant sa notoriété.

Deux autres catégories obtiennent des scores importants : la **notoriété de la marque** et la **notoriété des produits**. Ce sont des tweets qui parlent respectivement de la marque ou de l'un de leur produit et essaie de mettre ceux-ci en valeur. Il y a des marques comme **Nike** (17.55%), **Asics** (26.74%) ou **The North Face** (14.59%) qui optent plutôt pour la mise en avant de leurs produits par rapport à la marque. En termes de valeur absolue, **Puma** utilise ces stratégies autant que les autres marques mais en termes de pourcentage, ces tweets ne représentent qu'une petite partie de tout ce que la marque poste. En ce qui concerne la notoriété de la marque, cette catégorie vient majoritairement de tweets contenant le nom d'une célébrité ou d'un sportif de haut niveau. En effet, pour toutes les marques, plus de 57% des tweets de la stratégie ont été classifiés de telle façon car ils contenaient le nom d'une

personne connue et ce taux atteint même 97% pour **Salomon**. Cela, à l'exception de **Nike** qui n'a pas de tweets en rapport avec une célébrité et d'**Asics** dont le pourcentage s'élève à 46%.

The North Face (11.35%), **Adidas** (8.33%) et en moindre mesure **Salomon** (52 tweets) manifestent une stratégie de *responsabilité sociale* des entreprises qui est une stratégie qui émerge pour le moment avec tous les problèmes climatiques et sociaux que la Terre rencontre. Cependant, ce n'est pas une stratégie qui se démarque ici. Il est assez logique pour Salomon et The North Face de parler de responsabilité sociale car ce sont des marques connues pour leurs équipements de randonnée et cherchent donc davantage à proposer une image proche de la nature.

Les deux dernières stratégies, *promotion* et *saisonnnière* sont relativement peu présentes dans l'ensemble des tweets. Les promotions sont légèrement utilisées par **Asics** (6.65%), **Salomon** (4.82%) et en moindre mesure **Nike** (3.2%) tandis que la stratégie saisonnière est la plus utilisée par **New Balance** (8.33%), **Under Armour** (5.81%), **Salomon** (2.75%) et **The North Face** (2.7%).

La catégorie **Aucun** correspond à tous les tweets qui n'ont pas pu être classifiés dans les différentes stratégies. Les pourcentages de cette catégorie ne sont pas négligeables mais sont cependant plus faibles que lors d'une analyse de sentiments où plus de la moitié des tweets est classifiée comme neutre.

Ces résultats nous permettent d'émettre des observations intéressantes quant à l'axe autour duquel les marques ont choisi d'orienter leur stratégie.

4. Analyse des données de la communauté

Après avoir analysé comment se comportent les marques sur Twitter, nous pouvons dorénavant analyser quel est l'effet sur leur communauté, ou encore étudier le bouche-à-oreille (*word-of-mouth*) qui se propage sur Twitter à propos de ces neuf marques d'accessoires de sport.

Pour cela, nous allons tout d'abord, comme pour les tweets des pages, faire une première analyse descriptive de la base de données « *twitter_communauté* » qui contient tous les tweets dans lesquels le nom de la marque apparaît. On va ensuite faire une analyse de sentiments et finalement, une analyse des moments lors desquels les communautés postent sur Twitter. Ces moments seront considérés comme les moments lors desquels les communautés sont actives.

Comme nous avons pu le voir dans la FIGURE 02, la base de données intitulée « *twitter_communauté* » comporte des collections avec un nombre de tweets variant entre 25.800 et 942.000 pour l'année 2019, ce qui fait un total de plus de deux millions de tweets.

4.1. Première analyse descriptive

4.1.1. Analyse des mots les plus récurrents

À partir des tweets de la base de données « *twitter_communauté* », la fréquence de chaque mot a été calculée pour en ressortir les mots les plus fréquemment utilisés. Le TABLEAU 06 nous présente les résultats de cette analyse avec les 20 mots les plus récurrents pour chaque marque.

Le mot le plus tweeté pour chaque marque est évidemment le nom de cette marque. En effet, c'est sur ce critère que les tweets ont été sélectionnés.

On peut aussi voir que des mots tels que « *poshmark* », « *poshmarkapp* » ou « *shopmycloset* » font partie des vingt mots les plus récurrents pour sept des neuf marques analysées (**Adidas**, **New Balance**, **Nike**, **Puma**, **Reebok**, **The North Face** et **Under Armour**). Poshmark est un marché de commerce social en ligne où les utilisateurs, principalement américains, peuvent vendre et acheter des vêtements et accessoires, qu'ils soient neufs ou d'occasions. Poshmark est très actif sur Twitter, ce qui fait apparaître ces tweets dans une grande partie de notre base de données. De plus, l'application propose aux utilisateurs de partager des articles sur les réseaux sociaux, ils postent ainsi un tweet sur leur compte Twitter pour attirer les gens sur certains de leurs articles présents sur Poshmark. Un exemple de tweet est

« So good I had to share! Check out all the items I'm loving on @Poshmarkapp
#poshmark #fashion #style #shopmycloset #abas #ninewest #adidas:
<https://bnc.lt/focc/Jdol1QJrZQ> » @ozellobird7150

Ce genre de tweet est très répandu sur Twitter car la communauté Poshmark compte plus de 50 millions d'utilisateurs. C'est pourquoi, les mots cités précédemment apparaissent en premier dans les fréquences. Sur base des fréquences, les tweets en relation avec Poshmark représentent entre 5.8% pour **Puma** et 43.1% pour **The North Face** de tous les tweets incluant le nom de cette marque. La moyenne est de 15.7% pour les sept marques.

Dans le même mouvement, le mot « *ebay* » apparaît aussi fréquemment dans les mots fréquents. Ceci peut laisser suggérer que les articles de ces marques (**Reebok**, **The North Face** et **Under Armour**) sont aussi abondamment échangés sur le site d'enchères en ligne eBay et que des tweets émergent de cette plateforme.

Un détail à soulever est que le mot « *nike* » apparaît souvent dans les mots récurrents des autres marques (**Adidas**, **New Balance**, **Puma**, **Reebok**, **The North Face** et **Under Armour**). Cela peut venir du fait que, dans les tweets Poshmark, les articles de la marque **Nike** sont souvent cités en même temps que d'autres articles de la marque analysée. Cela peut aussi signifier que les utilisateurs comparent leur marque avec Nike, ce qui pourrait indiquer que Nike est perçue comme une marque de référence.

Le mot « *new* » est aussi présent pour toutes les marques, ce qui peut impliquer que ce sont des marques innovantes ayant souvent des nouveaux articles que les utilisateurs ont envie de partager. Cela peut aussi impliquer que les tweets à propos de ces marques font simplement référence à des articles achetés récemment par les utilisateurs mais pour ceux-ci, l'article est nouveau. Cependant, dans les deux cas, la nouveauté est un aspect important.

	Adidas		Asics		New Balance		Nike		Puma		Reebok		Salomon		The North Face		Under Armour	
	words	count	words	count	words	count	words	count	words	count	words	count	words	count	words	count	words	count
1	adidas	854489	asics	101795	new	174747	nike	945269	puma	375960	reebok	169650	salomon	35072	thenorthface	34628	armour	150021
2	check	85084	one	18275	balance	160054	air	154564	pumas	67656	check	31391	rondon	4296	check	16594	underarmour	82532
3	yeezy	84948	asicsza	17805	newbalance	45318	check	119389	check	31856	via	25108	running	3436	fashion	15500	check	68964
4	nike	79367	running	17387	shoes	25694	via	98731	new	30073	poshmark	20767	trail	2901	im	154606	poshmark	55280
5	boost	78136	springboks	15665	check	19042	poshmark	91829	via	26695	poshmarkapp	20690	shoes	2741	good	15373	poshmarkapp	55194
6	via	70669	shoes	15414	running	13772	poshmarkapp	91827	im	26580	shopmycloseset	20366	new	2357	style	15348	shopmycloseset	54447
7	new	63718	moveasone	15392	via	13593	shopmycloseset	89852	shoes	26039	shoes	17254	ski	2351	items	14981	via	51125
8	shoes	63537	strongertogether	15166	mens	12662	new	73658	poshmark	24204	new	13859	via	2143	poshmark	14968	added	32699
9	poshmarkapp	61594	logo	12219	poshmark	12620	im	69869	poshmarkapp	24156	added	13386	newcastle	1964	shopmycloset	14949	closet	31974
10	poshmark	61393	check	12006	poshmarkapp	12605	shoes	67178	shopmyclose	23542	mens	13166	amp	1799	poshmarkapp	14948	im	26447
11	shopmycloset	60041	gel	11595	shopmycloseset	12391	max	63857	nike	22247	closet	13019	speedcross	1731	share	14941	good	25088
12	size	57283	good	10569	nike	12199	size	63292	adidas	21310	size	12988	win	1656	loving	14926	fashion	23032
13	im	50873	jersey	10538	sneakers	11135	nikes	60705	black	20585	nike	12530	check	1642	north	4247	style	22977
14	x	43649	worn	10403	im	10488	added	58859	like	20560	im	12193	skis	1622	face	4192	items	22655
15	black	43490	new	9981	shoe	9960	closet	57472	x	17944	ebay	11422	mens	1608	jacket	2593	share	22573
16	style	40599	team	9848	size	8927	like	51618	good	17543	jersey	9867	sports	1519	nike	1760	loving	22127
17	added	39560	rugby	9663	x	8815	good	48353	sneakers	16398	adidas	9753	de	1476	new	1677	mens	18677
18	mens	38413	im	9438	added	8226	style	46339	get	15660	fashion	9382	one	1398	ebay	1639	new	17806
19	closet	38061	amp	9287	good	8144	jordan	45338	amp	15464	good	9341	mountain	1361	via	1626	nike	17436
20	like	35561	via	9098	closet	7754	fashion	44832	added	14778	style	9106	snow	1351	tnf	1137	ebay	17055

TABLEAU 06 – Mots les plus récurrents et leur nombre d'occurrence par marque pour la base de données « *twitter_communauté* »

4.1.2. Analyse du contenu des tweets populaires

Selon Hajikhani, Porras et Melkas (2017) [26], les top tweets jouent un rôle dans la perception de la marque par la communauté. Les top tweets sont détectés en fonction de l'impact qu'ils génèrent. Ici, les top tweets seront donc les tweets ayant le plus de « favorite ». Dans leur ouvrage, les auteurs affirment que les top tweets reflétant un sentiment négatif ont plus d'impact sur la communauté que les top tweets reflétant un sentiment positif. La communauté va donc plus facilement avoir un avis négatif sur la marque si les top tweets sont négatifs. Regardons l'émotion que les top tweets de nos différentes marques génèrent. Ici, le choix s'est porté sur un nombre de deux top tweets.

Voici les top tweets pour nos neuf marques :

"Willow Smith has teamed up with Adidas to create a 100% recycled shoe" @piscsfiles – 343.938 favorites

- "childish gambino airdropped a picture at coachella of his adidas collab and everyone who accepted got a pair of his unreleased shoes..." @dearapril – 294.783 favorites

"dudes will talk shit in a pair of asics. bro go for a run" @CaucasianJames – 27.072 favorites

- "Can't wait to see your faces again at the ASICS event happening March 16th! You just might be able to cop some custom designed sneakers done by yours truly ;) ;) #bethereorbesquare" @WE_THE_BOYZ – 8382 favorites

"Kawhi to off-camera teammate: \"You stupid. This ain't Burberry. ... This New Balance.\" " @ESPNBBA – 41.712 favorites

- "Man I don't care bout no sweden, I live in DC! I'm wearing New Balance, Solbiato and shit! I can't relate to no Swedish shit....." @geno_grigio – 23.944 favorites

"guy at the nike store asked me if any of their \"athletes\" helped me out today. calm down" @mariokartdwi – 277.102 favorites

- "We not 16 years old you not wearing my hoodies and keeping em. This Nike hoodie was \$100 foh I'll email your tweets to your job" @TheJayBurns – 185.971 favorites

"SS18 Fenty Puma by Rihanna " publieur inconnu – 108.527 favorites

- "Y'all have a weird attachment to black poverty. Rihanna's music is affordable, her makeup is affordable, her lingerie is affordable, her puma collaborations are affordable. She's wanted to do luxury and y'all got the audacity to be mad. It's not for EVERYDAY. Save up or shut up" @CoffeeCreamGirl – 72.161 favorites

"Off to the weekend @Reebok" @GalGadot – 26.484 favorites

- "On this day in 2003, LeBron signed a historic \$90M Nike deal at 18 years old.
 - □ Reebok handed him \$10M check to not talk to other brands; Said no
 - □ Chose Nike's legacy and signed for biggest initial deal ever
 - □ Built an empire worth \$1B+
 - □ Later got a lifetime contract"
- @BleacherReport – 26.368 favorites

"I changed snowboard to a new #salomon . All the clothing is #roxy . 場所: 舞子スノーリゾート
https://instagram.com/p/BuKk3DBuJ5/?utm_source=ig_twitter_share&igshid=1owir8fgigpg2"
@shimada_hrk1216 – 442 favorites

- "Marine Corps pack prep 2! This @salomon rain jacket was a great investment! It's so light that it barely exists but is 100% waterproof and breathable Bib on a belt so I can keep it outside, jacket or no" @jenrunswithdogs – 430 favorites

"We're proud to stand with @thenorthface in pledging to support our national parks when they need it most. Join us <http://brtsbe.es/nationalparks> #WeAreParks #GiveBackToNature #TrueForceofNature"
@BurtsBees – 17.488 favorites

- "Yesterday, we were disappointed to learn that @thenorthface and @LeoBurnett unethically manipulated Wikipedia. They have risked your trust in our mission for a short-lived consumer stunt. 1/" @Wikipedia – 4696 favorites

"a guy came up to me at the gym trying to be funny saying "you're not matching, you wearing under armour & nike. You cant do that haha" he lucky i didnt tell him his body aint matching too. Hitting all chest no legs" @darlachampagne – 82.704 favorites

- "When a chick who curved you walks in with a nigga that has Under Armour sneakers on. Weary face" @RegalCourtier7 – 45.306 favorites

Pour **Adidas**, les deux tweets sont positifs. On peut remarquer que dans le premier tweet, les sujets de recyclage et de nouveauté sont présents et dans le deuxième, c'est des chaussures inédites qui sont proposées. Le tweet le plus populaire pour **Asics** n'est pas très positif mais le deuxième l'est. Le premier post de **New Balance** est écrit par quelqu'un de connu ce qui fait du post quelque chose de positif tandis que le deuxième reflète un sentiment assez négatif qui n'est cependant pas visé sur la marque mais plutôt envers les médias. Les tweets populaires mentionnant **Nike** décrivent un sentiment penchant vers le négatif. Pour **Puma**, les deux top tweets sont en rapport avec Rihanna et véhiculent un sentiment positif. **Reebok** a un tweet qui fait la promotion de la marque et diffuse donc un sentiment positif tandis que son autre tweet est négatif pour la marque car c'est un tweet qui met Nike en valeur en dénigrant légèrement Reebok. Pour **Salomon**, les top tweets n'ont pas reçu beaucoup d'attention mais transportent cependant des messages positifs envers la marque. Le tweet le plus populaire pour **The North Face** est positif tandis que le deuxième est négatif car il parle du scandale de la marque avec Wikipédia (voir section 4.3.1). Finalement, **Under Armour** a des top tweets plutôt humoristiques dont un est positif et l'autre est plutôt négatif.

4.2. Analyse de sentiments

Comme il a été indiqué dans la section 3.2.2, il est intéressant de comparer les résultats des analyses de sentiments des pages avec les résultats d'analyses de sentiments identiques pour la communauté des marques. Pour ce faire, des analyses de sentiments ont été menées sur la base de données « *twitter_communauté* » et ce en utilisant le même principe que dans la section 3.2. Les trois mêmes ensembles d'entraînement ont aussi été repris.

La FIGURE 17 présente les résultats de l'analyse de sentiments en entrainant l'algorithme avec la 1^{ère} base de données.



FIGURE 17 – Graphiques représentant la proportion d'émotions par marque avec le 1^{er} ensemble de données

Comme pour la première analyse de sentiments de la section 3.2, la FIGURE 17 nous montre qu'il n'y a que cinq sentiments qui ressortent sur les treize présents dans l'ensemble d'entraînement.

Toutes les marques ont plus ou moins la même proportion de sentiments. Elles ont toutes entre 48% (pour **New Balance**) et 77% (pour **Adidas**) de sentiment *neutre*, suivies de 23% (pour **Adidas**) à 51% (pour **New Balance**) de sentiment *inquiète*. Elles ont aussi 3% ou moins des sentiments de *bonheur*, *tristesse* et *amour*. En suivant cette analyse de sentiments, c'est **Adidas** qui fait le moins penser les internautes à l'*inquiétude*, suivi par **The North Face** (27%) et **Under Armour** (32%). Par contre, **New Balance** (51%), **Nike** (50%) et **Salomon** (42%) y font penser le plus.

Les FIGURES 18 et 19 présentent les résultats des analyses de sentiments avec les deuxième et troisième ensembles d'entraînement, respectivement. Les proportions des sentiments relevés dans les tweets sont représentées pour chaque marque.

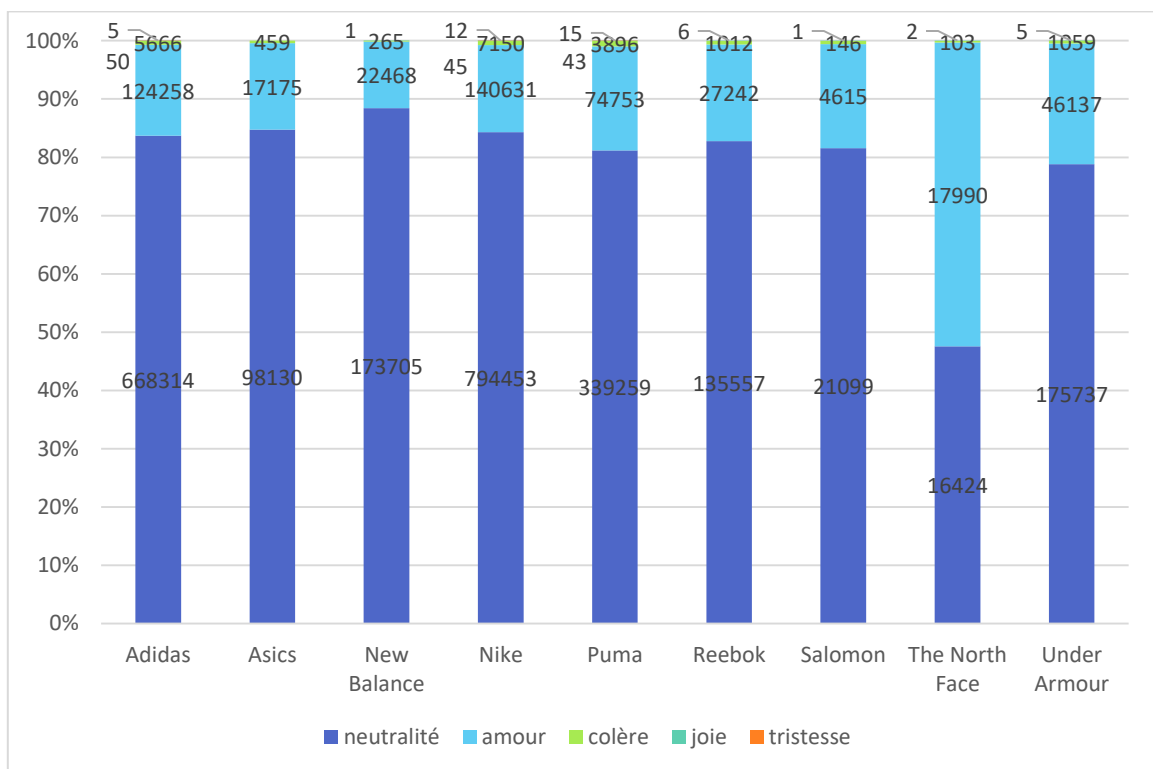


FIGURE 18 - Graphique représentant la proportion d'émotions par marque avec le 2^{ème} ensemble de données

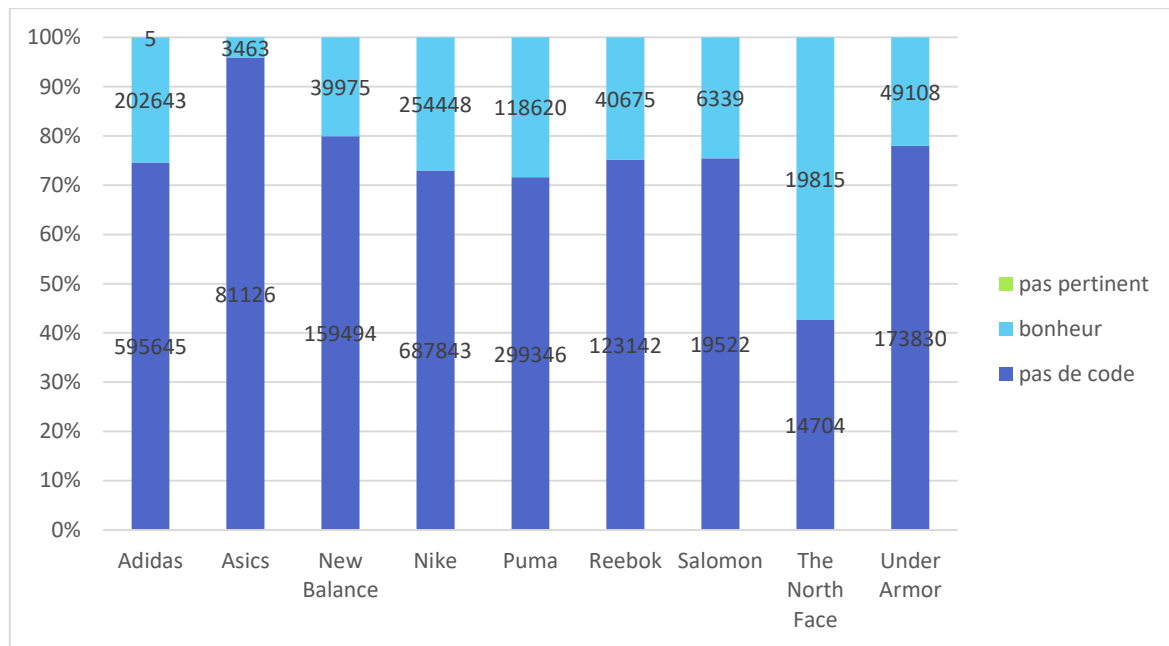


FIGURE 19 - Graphique représentant la proportion d'émotions par marque avec le 3^{ème} ensemble de données

Les résultats représentés dans les FIGURES 18 et 19 sont assez similaires avec **The North Face** qui a le plus de sentiments positifs, que ce soit le **bonheur** ou l'**amour** et les communautés des autres marques étant toutes environ au même niveau à l'exception d'**Asics** qui a très peu de sentiments de **bonheur** avec le 3^{ème} ensemble de données d'entraînement.

La **colère**, la **joie** et la **tristesse** apparaissent avec le 2^{ème} ensemble. La **colère** est présente en moindre quantité pour toutes les marques à l'exception de **New Balance** et la **tristesse** est présente pour **Adidas**, **Nike** et **Puma**, les collections comprenant le plus de tweets. Quant à la **joie**, elle n'est pas distinguable sur le graphique, ce qui signifie que sa proportion est infime et non pertinente.

Pour une meilleure analyse, tous ces résultats seront à comparer avec ceux de la section 3.2.2.

4.3. Analyse du moment des publications

Comme dans la section 3.3, une analyse a été réalisée sur le moment où les posts ont été publiés sur Twitter pour chaque communauté des marques. Cependant, le but n'est pas tout à fait le même. Dans la première analyse, le but était de comprendre les différentes stratégies mises en place par les marques. Ici, le but est plutôt de détecter quand la communauté publie des posts et donc à quel moment est-elle la plus active. De plus, nous verrons que certains événements peuvent impacter les communautés.

4.3.1. Semaine et événements marquants

La FIGURE 20 présente un graphique qui reprend une courbe pour chaque marque indiquant le pourcentage de tweets postés pendant la semaine par rapport à l'année entière.

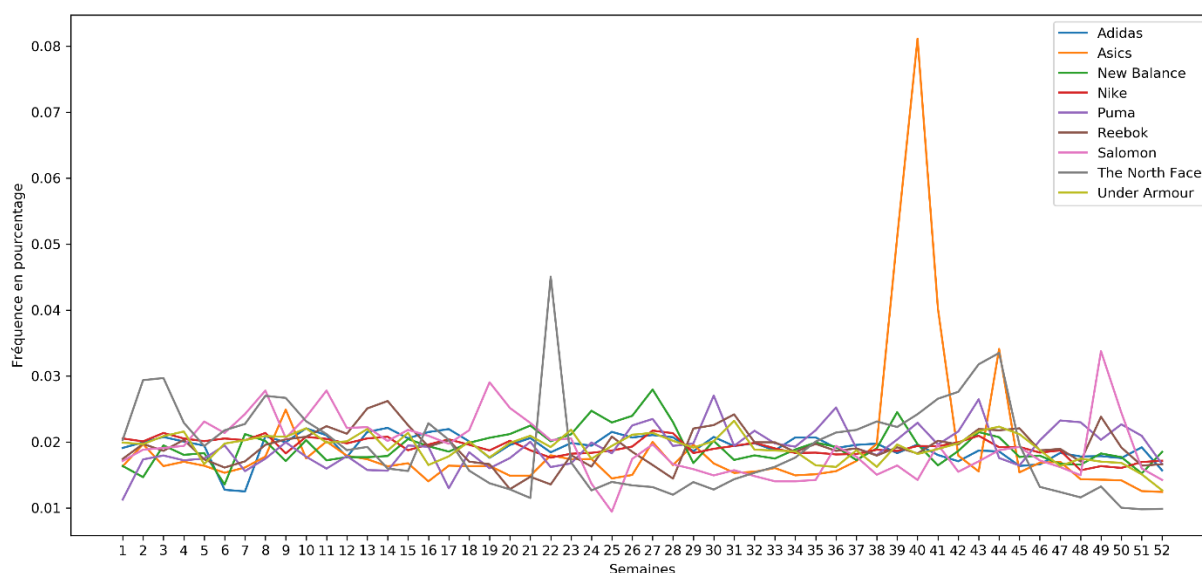


FIGURE 20 – Graphique représentant le nombre de tweets en pourcentage par communauté de marque et en fonction de la semaine

La FIGURE 21 présente les mêmes résultats que la figure précédente excepté la communauté de la marque Asics qui a été supprimée par souci de visibilité.

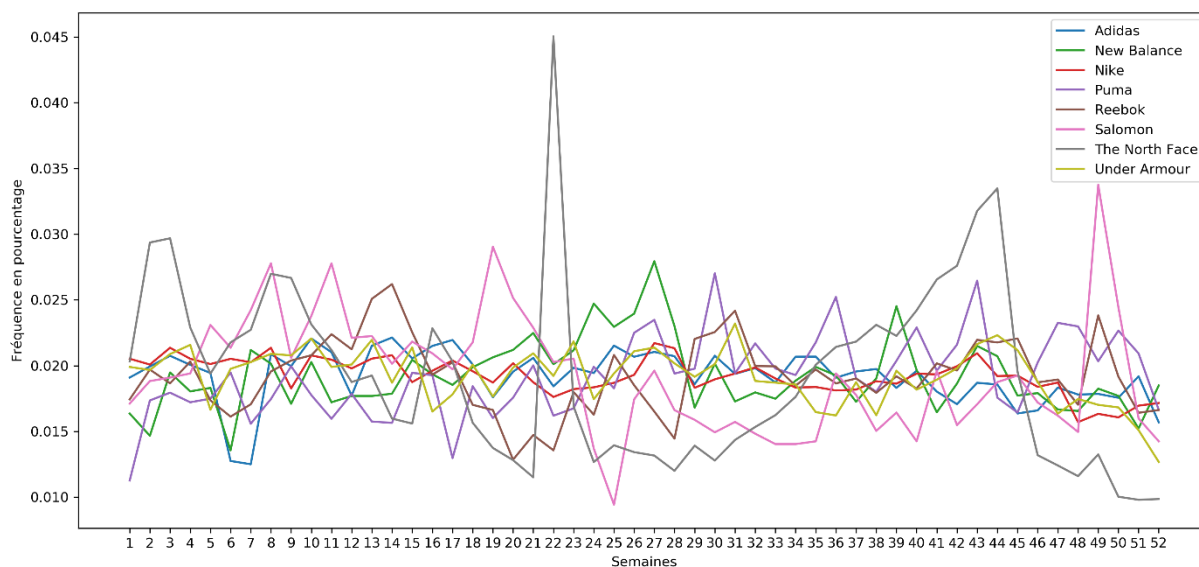


FIGURE 21 – Graphique représentant le nombre de tweets en pourcentage par communauté de marque (excepté Asics) et en fonction de la semaine

Au vu des FIGURES 20 et 21, il n’y a pas de tendance à la hausse ou à la baisse au cours de l’année. Le nombre de tweets suit une constante avec des pics de temps en temps.

Le pic le plus notable est celui d’**Asics** lors de la 40^{ème} semaine, c’est-à-dire du 30 septembre au 6 octobre 2019. Lors de cette semaine, la communauté a publié quatre fois plus qu’à son habitude. Ce pic est dû à la page d’Asics sud-africaine qui a lancé un concours lors duquel les utilisateurs devaient deviner de quand datait une photo des Springboks, l’équipe de rugby sud-africaine dont la marque est équipementier. Cette campagne a récolté environ 6500 réponses et a donc porté ses fruits.

Le deuxième pic revient à **The North Face** lors de la 22^{ème} semaine, du 27 mai au 2 juin 2019. C’est le moment lors duquel la marque a dû faire face à un scandale en relation avec le site de l’encyclopédie collective Wikipédia. En effet, Wikipédia a dénoncé la marque pour avoir, sans autorisation, changer les photos de paysage apparaissant sur l’encyclopédie. Ces nouvelles photos avaient toutes un point commun, elles faisaient apparaître un article de la marque de sport. The North Face avait donc trouvé un moyen de faire de la publicité gratuitement mais Wikipédia ne s’est pas laissé faire. La courbe de The North Face présente aussi deux autres pics, aux 2^{ème} et 3^{ème} semaines et à la semaine 44. Ce deuxième pic est intéressant car il croît pendant plus de dix semaines pour redescendre drastiquement juste après la 44^{ème} semaine. Les tweets de la semaine à l’apogée viennent du marché de commerce social Poshmark. On peut imaginer que le site a tenu les utilisateurs en haleine jusqu’à cette semaine d’apogée et leur attention a par la suite été perdue. Le premier était aussi dû aux tweets en relation avec Poshmark.

Salomon possède le 4^{ème} plus grand pic lors de la 49^{ème} semaine. Il correspond à un concours lancé par la marque et ayant comme prix une paire de ski. Cette marque possède aussi un pic négatif lors de la 25^{ème} semaine. Cela pourrait correspondre à une baisse d'intérêt de la part des consommateurs.

Comme dans la section 3.3.1, regardons si les plus grands événements sportifs (voir ANNEXE 03) ont un impact sur le contenu des publications de la communauté. L'événement qui a le plus d'impact est la *ligue de basket-ball nord-américaine*, la NBA. En effet, le mot « *nba* » revient plus de 13.500 fois dans les tweets de la communauté de **Nike**, plus de 4000 fois pour **Adidas** et entre 1200 et 1700 pour **Reebok**, **Puma**, **Under Armour** et **New Balance**. Il y a aussi quatre autres ligues professionnelles majeures dans le sport américain, à savoir la NFL pour le football américain, la MLB pour le baseball et la NHL pour le hockey sur glace. Les deux premières ont aussi un impact sur les communautés des marques. L'impact le plus grand est celui de la NFL sur **Reebok** car le mot « *nfl* » apparaît plus de 7500, c'est le 22^{ème} mot le plus fréquent pour cette communauté. Il apparaît aussi pour **Nike**, **Under Armour** et **Adidas** et dans une moindre mesure pour **The North Face**, **Puma** et **New Balance**. La finale de ce championnat qui est un événement bien connu de tous, le *Super Bowl*, se montre aussi dans les communautés de **Reebok**, **Nike** et **Adidas**. **Reebok** connaît également un impact de la MLB mais moins important avec le mot « *mlb* » s'affichant 957 fois. Ce mot s'affiche aussi pour **Under Armour**, **Nike**, **New Balance** et **Adidas**.

En ce qui concerne *Wimbledon*, le mot est présent pour **New Balance** mais seulement 101 fois, ce qui le place à la 2346^{ème} position des mots les plus fréquents. Un autre championnat de tennis est aussi présent dans les communautés, il s'agit de l'*US Open* organisé à New York. Il apparaît pour **New Balance** et **Nike**. Le *championnat d'athlétisme* aux Etats-Unis se démarque aussi dans les deux mêmes communautés.

La *coupe du monde de football féminin* est présente en moindre mesure dans les communautés d'**Adidas** (215 fois), **Nike** (352 fois) et **Puma** (298 fois). La *Ligue des Champions* qui est une compétition de football entre les clubs européens apparaît aussi sur Twitter et ce en lien avec **New Balance**, **Adidas** et **Nike**. Le *marathon de Boston* se révèle dans deux communautés, à savoir **New Balance** et **Asics** mais aussi dans une moindre mesure avec 43 et 25 apparitions respectivement. La communication d'Asics envers cet événement a donc potentiellement fonctionné mais pas pour **Salomon**. La *coupe du monde de rugby* se manifeste quant à elle dans les tweets parlant d'**Asics** et de **Puma**. Finalement, les *CrossFit Games* apparaissent 624 fois dans la communauté de **Reebok** qui avait misé une grande partie de sa communication sur cet événement.

Salomon fait référence à un bon nombre d'événements dans sa communication comme vu dans la section 3.3.1 mais ces événements ne se distinguent pas réellement dans sa communauté. En effet,

le *Tahoe Rim Trail* apparait seulement 20 fois et le *Pikes Peak Trail* 4 fois. La série de courses *Golden Trail Series* est légèrement plus présente avec un minimum de 96 apparitions. Finalement, le *SalomonWMN camp* est mentionné 42 fois.

En ce qui concerne les pics repérés dans la FIGURE 12 (de la section 3.3.1), la remise d'un Oscar au film « Free Solo » sponsorisé par **The North Face** est présente en petite quantité dans la communauté de la marque avec 95 fois l'association de mot « alexhonnold ». Par contre, le #CreatorsUnite n'est vraiment pas présent en proportion du nombre de tweets pour **Adidas**. En effet, sur les 798.527 tweets, le hashtag s'y retrouve 87 fois.

4.3.2. Jour de la semaine

La FIGURE 22 présente les résultats de l'analyse des moments lors desquels la communauté publie sur Twitter en fonction des jours de la semaine.

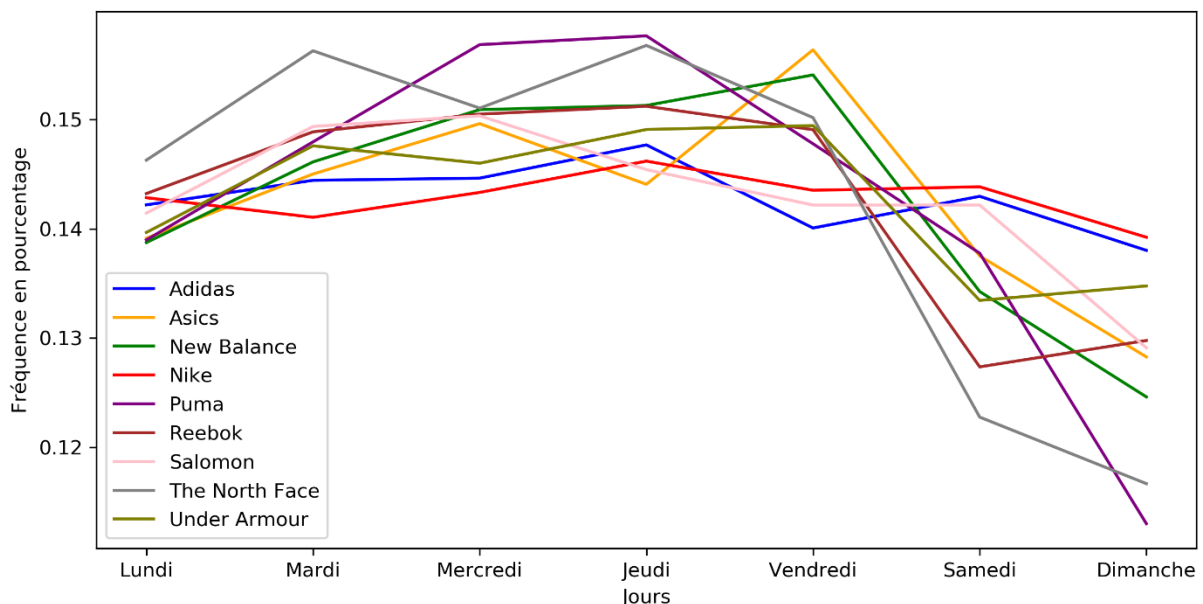


FIGURE 22 – Graphique représentant le nombre de tweets en pourcentage par communauté de marque et en fonction du jour de la semaine

Dans la FIGURE 22, la tendance est nettement dessinée. Pendant les jours de la semaine, le nombre de tweets est assez constant excepté pour le lundi où il y a un peu moins de tweets. Pendant le week-end, la tendance est à la baisse avec une diminution le samedi et une autre le dimanche. Ceci laisse suggérer que les utilisateurs de Twitter sont moins présents sur le réseau social pendant le week-end et qu'il est donc préférable pour une marque de ne pas poster pendant cette période car sa publication risquerait de ne pas être vue en suffisance.

4.3.3. Heure

La FIGURE 23 présente les résultats de l'analyse des moments lors desquels la communauté publie sur Twitter en fonction de l'heure pendant la journée. L'heure est aussi en UTC.

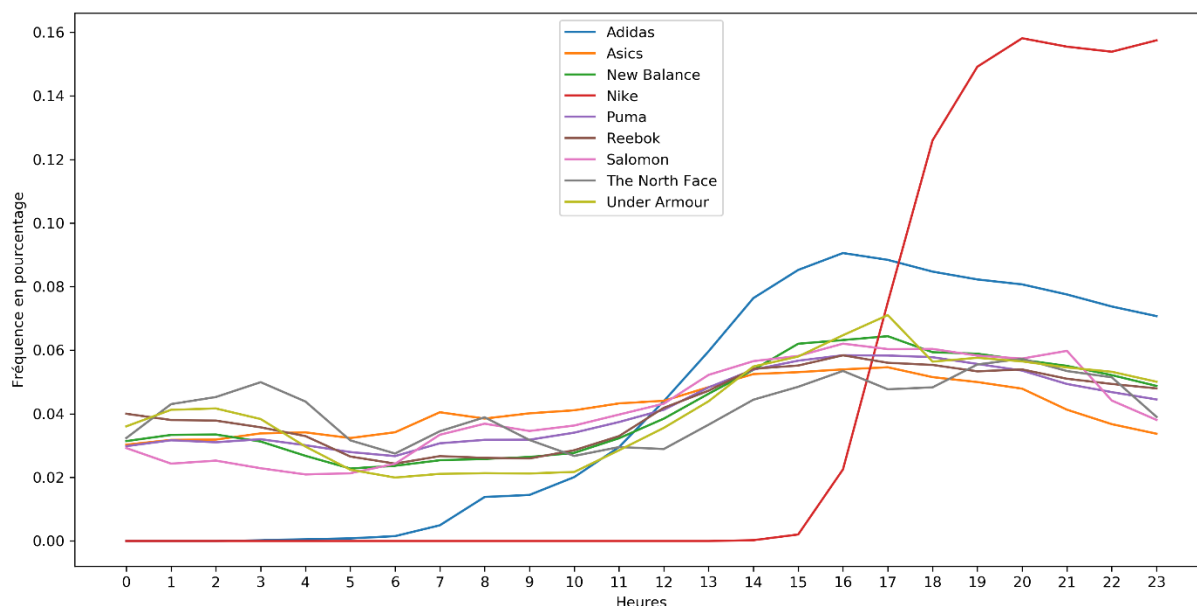


FIGURE 23 – Graphique représentant le nombre de tweets en pourcentage par communauté de marque et en fonction de l'heure

La tendance présente dans la FIGURE 23 est assez nette. Il y a une augmentation du nombre de tweets à partir de 12h UTC qui atteint son pic vers 17h pour diminuer jusqu'à 23h. Si l'on continue avec la journée suivante, le nombre de tweets diminue même jusqu'à 6h UTC.

Ici, le biais dans les données des marques **Nike** et **Adidas** présenté dans la section 1.4 est visible. En effet, l'algorithme de collecte de données n'a pris que les premiers tweets par jour pour chaque marque et l'on peut voir sur ce graphique que l'algorithme commençait par la fin de la journée. En partant de 23h, la limite de tweets était toujours atteinte avant 14h pour Nike et avant 5h pour Adidas. Leurs deux courbes ne peuvent donc pas être étudiées ici.

5. Analyse transversale

Après avoir analysé le contenu des pages de nos neuf marques de sport et des tweets parlant de ces marques, il est temps de mettre en parallèle ces deux analyses.

5.1. Remarques générales

Regardons tout d'abord le nombre de tweets postés par les marques et le nombre de tweets venant de leurs communautés respectives et ajoutons-y une information sur la grandeur de la marque pour améliorer cette analyse. Pour cela, la FIGURE 24 présente le nombre de tweets pour la communauté de chaque marque ainsi que leur chiffre d'affaire qui sera une variable indicatrice de la grandeur de la marque ou de sa notoriété.

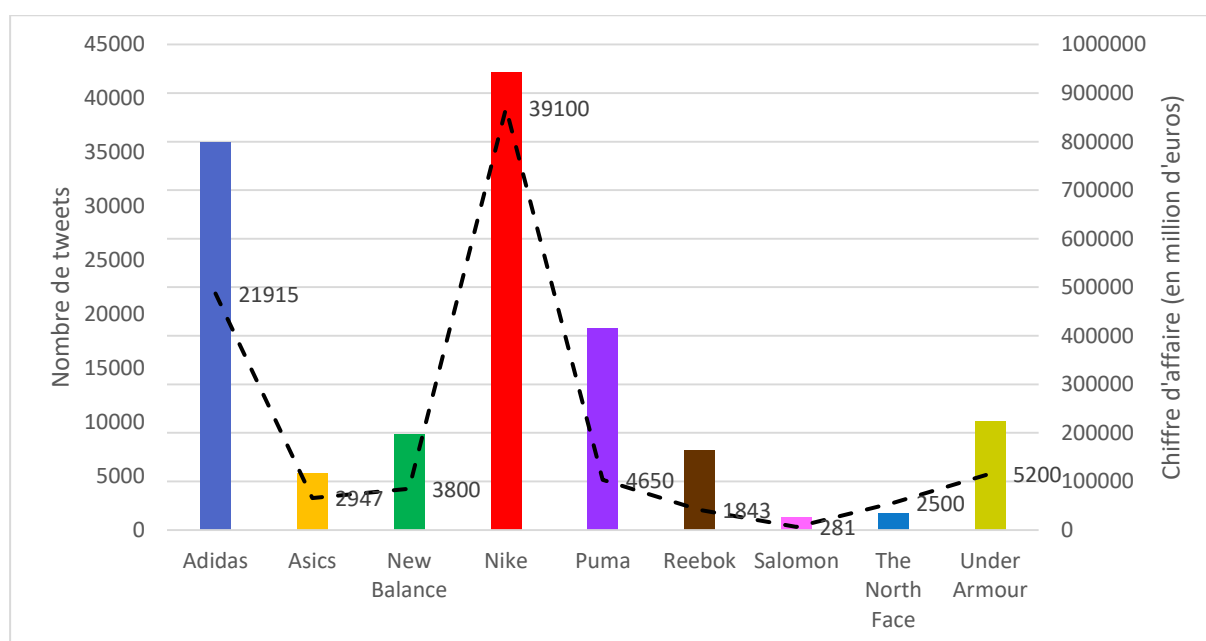


FIGURE 24 – Graphique représentant le nombre de tweets et le chiffre d'affaire par marque⁷

Dans cette figure, on remarque que le nombre de tweets dans la communauté est tout à fait proportionnel au chiffre d'affaire de la marque. Une marque dont on parle plus sur Twitter aura donc une plus grande notoriété qu'une marque dont on parle peu comme Salomon.

Dans les sections précédentes, nous avons pu remarquer que c'est **Adidas** qui poste le moins sur sa page Twitter. Cependant, on a aussi pu noter qu'elle fait partie des marques dont la collection pour sa communauté est la plus grosse. De plus, elle fait aussi partie des marques qui ont le plus de followers. Adidas poste donc assez de tweets pour garder l'attention de ses utilisateurs. On pourrait croire que le fait d'avoir déjà une certaine notoriété implique d'utiliser une stratégie épurée.

⁷ Le chiffre d'affaire est celui de l'année 2018 sauf pour Reebok où il est de 2017

Cependant, **Nike** qui a encore une plus grande notoriété n'a pas choisi cette stratégie. En effet, Nike poste dix fois plus que son concurrent Adidas sur Twitter et a aussi un chiffre d'affaire plus imposant. Une grande majorité des posts est constitué de réponses à des utilisateurs. Ce sont donc des publications qui ne touchent pas beaucoup de monde. Cependant, Nike poste aussi des publications avec un contenu direct destinés à un plus grand nombre de followers. Elle a simplement décidé de répondre à ses consommateurs par surcroit. Chaque marque adopte donc une stratégie bien spécifique à son entreprise et à ses objectifs.

Pour **Salomon** qui épouse la même stratégie que Nike, il s'agit de l'inverse d'Adidas. La marque poste beaucoup sur sa page mais sa communauté est beaucoup plus petite et son nombre de followers plus faible. Si la marque poste autant, c'est peut-être justement pour attirer plus de followers et avoir plus d'utilisateurs parlant d'elle dans leurs tweets. En effet, Salomon est une entreprise en pleine expansion. Elle doit donc faire des efforts pour essayer de s'étendre. Il ne faut pas oublier qu'une grande partie des tweets postés par la marque vient de réponses à ses utilisateurs. Cela signifie que la marque se veut proche de ses consommateurs.

Une motivation pour **Adidas** de ne pas poster beaucoup est éventuellement que la qualité prime sur la quantité. En effet, l'étude de Hajikhani, Porras et Melkas (2017) [26] révèle que, pour les publications sur les réseaux sociaux, c'est la qualité qui est plus importante que la quantité. Cela amplifie l'importance du contenu représentant les entreprises sur les réseaux et insiste sur les méthodes innovantes de production et de délivrance. Adidas poste moins sur Twitter mais le fond de ses tweets est peut-être plus étudié. De plus, toujours selon cette même étude, un contenu de qualité, c'est-à-dire un contenu qui plait aux utilisateurs, est basé sur un aspect plus humain plutôt que sur de la publicité pure. On peut voir cette tendance dans l'analyse des top tweets. En effet, ces tweets parlent de recyclage, de pauvreté, de parc national, reprennent des phrases de célébrités dites spontanément ou font de l'humour. Nous pouvons donc confirmer l'hypothèse amenée par Hajikhani et al et conseiller aux marques d'adopter de tels sujets lorsqu'elles composent leurs tweets ou mêmes toutes autres publications sur les réseaux sociaux.

Grâce à l'étude des mots les plus fréquents dans les communautés et à l'analyse du moment des publications, nous avons pu dégager l'effet qu'ont certains scandales ou certaines campagnes sur les utilisateurs. En effet, nous avons pu remarquer dans la section 3.3.1 que **The North Face** misait une partie de sa communication sur l'Oscar du meilleur documentaire qu'a remporté le film Free Solo parlant de l'ascension faite par Alex Honnold, un athlète de l'équipe. En regardant dans les mots les plus fréquents de la communauté, nous pouvons voir que les mots « freesolo » et « alexhonnold » sont les 821^{ème} (le mot est cité 54 fois) et 431^{ème} (le mot est cité 96 fois) respectivement. Nous pouvons

aussi voir se dessiner un léger pic à la date de la remise de l'Oscar lors de la neuvième semaine sur la FIGURE 21 (section 4.3.1.). Nous pouvons ainsi affirmer que la communication a été efficace car elle a eu un effet sur les utilisateurs même si celui-ci est modéré. Un autre évènement a eu un impact sur la marque mais ce dernier fait moins la publicité de celle-ci. En effet, il s'agit du scandale avec Wikipédia. Le plus grand pic pour la marque dans la FIGURE 21 est d'ailleurs dû à cette affaire. De plus, un des top tweets y est également lié. Ceci confirme l'hypothèse de Hajikhani, Porras et Melkas (2017) [26] qui affirme que les éléments négatifs ont plus d'impact que les éléments positifs. Au vu de ces résultats, il est vraiment important pour les entreprises d'essayer de se garder à distance des scandales car ils sont vite répandus sur les réseaux sociaux. Cependant, au vu de l'analyse de sentiments de The North Face, nous avons pu remarquer que c'était la marque qui avait le plus de positivité dans sa communauté. Le scandale a certainement joué un impact mais il n'a peut-être pas été aussi négatif qu'on pourrait le penser. Finalement, on a aussi pu remarquer qu'un concours réalisé par la marque **Asics** a porté ses fruits car il a récolté environ 6500 tweets dans la communauté de la marque. Nous pouvons ainsi noter qu'une analyse de ce qui est dit à propos des marques sur les réseaux sociaux est une bonne technique pour évaluer l'efficacité d'une campagne de marketing.

5.2. Analyse des sentiments

Comparons maintenant les analyses de sentiments menées pour nos deux bases de données afin de comprendre si les sentiments que les marques veulent transmettre se retrouvent dans leur communauté.

Avec le premier ensemble d'entraînement, on peut voir que la majorité des marques (**Adidas**, **Asics**, **Reebok**, **Salomon**, **The North Face** et **Under Armour**) ont moins de sentiments *neutres* sur leur page comparé à leur communauté mais ont un pourcentage plus élevé de sentiments de *bonheur* même si la proportion de celui-ci est assez infime, ne dépassant pas les 13%. **New Balance** et **Nike** ont aussi plus de sentiments de *bonheur* sur leur page que dans leur communauté mais le sentiment de *neutralité* est aussi supérieur. Ceci démontre que ces marques parlent plus positivement d'elle-même que ce que ne fait leur communauté. Cependant, il y a une marque pour laquelle c'est le contraire. Il s'agit de Puma. Cette marque est aussi celle ayant le moins de sentiments de *bonheur* venant de sa page, on peut donc suggérer que, même si la page véhicule très peu de sentiments de *bonheur*, sa communauté est satisfaite et exprime ce sentiment de manière plus considérable. De plus, avec **Under Armour** et **Salomon**, Puma fait partie des marques dont la communauté poste le plus de tweets faisant référence au *bonheur*.

Avec les deuxième et troisième ensembles d'entraînement qui donnent des résultats assez similaires, on peut voir que c'est aussi le sentiment *neutre* qui prime partout. Cependant, pour **Adidas**, **Asics**, **New Balance**, **Nike**, **Reebok** et **Salomon**, il y a dans tous les cas moins de sentiments de *neutralité* sur leur page et plus de sentiments *positifs* en comparaison avec leur communauté respective. Les marques véhiculent donc un sentiment plus *positif* via leur page que ce que n'exprime leur communauté. Les pages de ces marques sont celles qui expriment la proportion la plus élevée de sentiments *positifs* et c'est pourquoi nous arrivons à ce résultat. De plus, leur communauté ont le plus de sentiments *neutres*. Les marques essaient de faire passer un sentiment *positif* mais ce n'est pas toujours bien reçu par leur communauté. Pour **The North Face**, la tendance est à l'inverse car la page de la marque possède beaucoup plus de sentiments *neutres* que les autres marques. C'est cependant une bonne chose pour la marque qui obtient le meilleur score de sentiments *positifs* dans sa communauté malgré le fait qu'elle ait dû subir le scandale Wikipédia. Finalement, pour **Puma** et **Under Armour**, les résultats ne sont pas tout à fait tranchés donc aucune conclusion ne peut être tirée.

En croisant ces résultats, on peut affirmer que pour toutes les marques, c'est le sentiment de *neutralité* qui prédomine. Pour **Adidas**, **Asics**, **New Balance**, **Nike**, **Reebok**, **Salomon** et **Under Armour**, il y a plus de sentiments *positifs* diffusés par leur page qu'ils n'en présentent dans leur communauté. C'est un résultat qu'on pouvait espérer car les marques essayent du mieux possible de propager une

image positive d'elle-même et elles le font via leurs tweets qui reflètent donc un sentiment **positif**. Les utilisateurs écrivant des posts à propos de ces marques font souvent moins usage de ce sentiment **positif**. Il y a cependant deux marques qui dérogent à la règle, **Puma** et **The North Face**. Dans deux des analyses, les marques ont plus de sentiments **positifs** dans leur communauté que sur leur page. Pour The North Face, cela est dû au fait que la communauté montre un sentiment plus **positif** à l'égard de la marque. C'est donc un avantage pour The North Face qui est mieux vue par les utilisateurs. Quant à Puma, cela est dû au fait que la marque poste moins de choses **positives** sur sa marque. Les tweets de la page de Puma ont en effet la plus grande partie de réponses en comparaison avec toutes les autres marques (voir FIGURE 05 de la section 3.1.1) et les réponses étant certainement moins étudiées pour véhiculer un sentiment positif.

C'est certainement aussi pour cette même raison, le fait que la majorité des tweets soient en réalité des réponses aux utilisateurs, que le sentiment de neutralité est le plus prépondérant pour toutes les marques.

5.3. Analyse des moments opportuns

Dans cette sous-section, nous allons comparer les résultats de l'analyse de la stratégie de fréquence de la section 3.3 et de l'analyse du moment des posts de la section 4.3. L'objectif est de comprendre à quel moment il est préférable pour les marques de publier un post en fonction des réactions « favorite » mais aussi en fonction du moment où leur communauté publie le plus.

En fonction des semaines, il n'y a pas spécialement de meilleurs moments pour poster du contenu. En effet, aucune tendance ne se dessine sur les graphiques des FIGURES 21 et 22. Seuls des pics apparaissent et sont généralement dû à des événements spéciaux.

En fonction du jour de la semaine, c'est les posts du *jeudi* qui obtiennent le plus d'attention, c'est-à-dire de mention « favorite ». Le *jeudi*, les posts des différentes communautés sont aussi dans leur zone maximale. C'est donc une bonne chose pour les marques car cela signifie que les communautés sont actives et qu'il y a plus de chance qu'elles voient le tweet qui vient d'être posté par la marque. Les publications du *week-end* obtiennent aussi plus d'attention que les autres. Cependant, c'est lors de ces deux jours que les communautés sont les moins actives, le nombre de tweets postés est le plus faible le week-end mais est le plus élevé entre mardi et vendredi. Par conséquent, ce n'est pas parce que les communautés ne publient pas sur Twitter qu'elles ne sont pas actives sur les réseaux et posent des mentions sur les posts des marques. Si l'on se fie à l'analyse d'Elizabeth Arens (2020) [24], c'est le *dimanche* suivi du *mercredi*, qu'il est plus opportun pour les marques de biens de consommations de poster sur Twitter. Le nombre de « favorite » confirme cela à l'exception du *jeudi* qui est moins présent dans l'étude d'Elizabeth Arens. Si l'on rassemble donc ces deux résultats, les jours lors desquels il est plus favorable pour les marques de poster sont le *mercredi*, le *jeudi* et le *dimanche*. Chaque marque a adopté une stratégie un peu différente en ce qui concerne le jour où elle poste un tweet et aucune ne semble avoir privilégié ces trois jours même si une légère augmentation est souvent dénotée le dimanche par rapport au samedi alors que ce sont deux jours fériés. Il y a bien **Reebok** qui privilégie le *jeudi* mais n'est pas très présente les deux autres jours. Un gain est donc à faire pour les marques en ce qui concerne leurs jours de prédilection pour poster une publication.

Finalement en ce qui concerne l'heure, c'est entre 14h et 17h UTC que le nombre de « favorite » est le plus important et c'est aussi à cet horaire que nos neuf marques publient le plus sur Twitter. Ces trois heures font aussi partie de celles pendant lesquelles la communauté des marques est la plus active car elles postent le plus de posts entre 12 et 21h UTC. Dans l'étude d'Elizabeth Arens, l'horaire le plus efficace est entre 10h et 15h. Si l'on considère que les posts sont à l'heure européenne, la translation de l'horaire de l'étude donne 12h – 17h UTC. Tout ceci correspond et implique donc que les marques savent à quel moment il est préférable de poster sur Twitter.

5.4. Stratégie de communication

Finalement, regardons ce que le croisement entre l'analyse des stratégies de communication et tous les autres éléments de notre étude peut nous apprendre sur les différentes marques.

Au vu du nombre élevé de réponses dans les tweets, nous pouvions nous attendre à ce que la plus grande proportion de tweets tombe dans la stratégie **service client et engagement**. Ces pourcentages sont d'ailleurs proportionnels à la quantité de réponses présentes pour chaque marque comme on peut le voir dans la FIGURE 04 de la section 3.1.1.

Le taux de stratégie de **responsabilité sociale** est quant à lui l'un des plus bas. Les entreprises ne concentrent pas encore leur stratégie marketing autour de cela. Cependant, comme nous avons pu le voir dans la section 4.1.2, le tweet venant de la communauté d'**Adidas** et ayant le plus de « favorite » parle de recyclage, c'est donc quelque chose qui attire les consommateurs. Adidas fait d'ailleurs mieux que ses concurrents, à l'exception de **The North Face**, en ce qui concerne la responsabilité sociale. C'est une stratégie qui a du futur et qui pourrait être mieux exploitée par ces entreprises, surtout en ces temps de lutte contre le changement climatique. Ce même tweet parlant de recyclage parle aussi d'une célébrité, à savoir Willow Smith, la fille du célèbre acteur Will Smith. Inclure des personnes populaires est donc aussi une stratégie gagnante. Les deux top tweets de **Puma** mentionnent d'ailleurs la chanteuse à la renommée mondiale, Rihanna. Cette stratégie prend part dans celle intitulée « **Notoriété de la marque** ». Comme vu dans la section 3.4, c'est même les tweets mentionnant ces personnes connues qui constituent environ 80% de la catégorie. Sponsoriser et faire des partenariats avec des célébrités ou des sportifs de haut niveau sont donc des stratégies qui fonctionnent pour les marques de sport et cela, les entreprises l'ont bien compris et l'appliquent dans leur communication sur Twitter. **Nike** est cependant une exception car elle ne mentionne pas ce genre de personnes.

En ce qui concerne la stratégie **saisonnnière**, elle n'est pas non plus très présente dans les pages de nos neuf marques. Cependant, les marques de sport organisent, sponsorisent ou parrainent assez régulièrement des événements sportifs ou des athlètes qui participent à ces événements. Comme nous avons pu le voir dans la section 3.3.1, certaines marques comme **Salomon** aiment communiquer à propos de ces événements. Cependant, lorsque l'on regarde l'effet que ce genre de communication a sur la communauté, nous pouvons remarquer qu'il est assez limité. C'est peut-être pour cette raison que les marques utilisent moins cette stratégie sur les réseaux sociaux. Attention, ce n'est pas pour autant qu'elles ne l'utilisent pas dans d'autres contextes.

La stratégie de **promotion** est elle aussi peu utilisée sur les réseaux sociaux. Comme le disait Hajikhani, Porras et Melkas (2017) [26], ce n'est pas le genre de contenus que les utilisateurs apprécient le plus sur les réseaux sociaux et nos différentes marques semblent l'avoir compris.

Finalement, en ce qui concerne la **notoriété des produits**, c'est la deuxième stratégie la plus utilisée dans notre base de données. Les marques aiment mettre en valeur leurs produits et ce sont généralement de nouveaux produits. En effet, l'innovation est quelque chose d'important pour les marques de sport. La différenciation l'est aussi et c'est par leurs produits que les marques réalisent cela, c'est pourquoi ce genre de communication est prisé par les entreprises de vêtements ou accessoires.

6. Conclusion

Ce mémoire avait pour ambition de répondre à la question de départ « Comment se comportent les marques de sport sur Twitter et quel est l'effet sur leur communauté ? ». Pour cela, nous avons mis en place des éléments d'analyse pour nos bases de données qui ont récolté les tweets des pages et de leur communauté. Ces éléments sont l'analyse du profil général des données, les analyses de sentiments, les analyses du moment des publications et finalement, l'analyse de la stratégie de communication pour les données des pages de nos neuf marques.

Grâce à ces différentes analyses, nous avons pu produire des résultats intéressants. Nous avons vu que de manière générale, les marques adoptent différentes stratégies de fréquence car certaines postent beaucoup plus que d'autres et ce parce qu'elles répondent à leurs utilisateurs. Ensuite, nous avons aussi pu remarquer que les marques de sport axent une grande partie de leur communication sur leur sponsoring que ce soit avec des équipes de football, avec des athlètes ou pour des événements. Cette communication a souvent un impact sur la communauté mais il est assez léger en comparaison avec les proportions utilisées pour la communication. Au niveau de l'analyse de sentiments, nous avons remarqué qu'à l'exception de **The North Face**, les marques n'arrivent pas à transmettre autant de sentiments positifs à leur communauté en analogie avec ce qu'elles véhiculent dans leurs publications. Cependant, les communautés obtiennent tout de même un score de sentiments positifs encourageant. En ce qui concerne les moments les plus opportuns pour poster des tweets, ce sont les mercredis, jeudis et dimanches et ce entre 14h et 17h UTC. Les marques semblent avoir compris pour l'heure la plus favorable mais leur stratégie en fonction du jour de la semaine ne reflète pas la même chose bien qu'elles augmentent leurs tweets le dimanche alors que c'est un jour férié. Finalement, la stratégie de communication qui ressort le plus est celle du **service client et engagement** et ce car la majorité des tweets sont des réponses. La **notoriété de la marque** et **des produits** sont aussi assez prisées par nos neuf organisations. Nous avons aussi remarqué que la stratégie de **responsabilité sociale** n'était pas encore utilisée à sa juste valeur.

Étant donné le nombre d'éléments à prendre en compte pour définir la stratégie globale d'une marque, il est normal que les marques adoptent leur propre stratégie et se comportent donc de manière assez différente. Il y a tout de même des ressemblances dans le fond comme l'idée de partager leur sponsoring avec des célébrités et des événements ou de poster lors des moments les plus actifs d'une journée. Quant aux communautés, elles réagissent de manière plus ou moins intense à cette communication en fonction de son type mais tout de même en moindre mesure au vu de la manière dont elles régissent lorsqu'un scandale explose.

Notre étude a donc répondu à la question de base et a proposé différentes méthodes d'analyse de la communication des organisations sur les réseaux sociaux. Il pourrait être intéressant et enrichissant de mettre en place cette même étude avec un autre type de marques ou même sur un autre réseau social.

Cependant, notre travail compte quand même quelques limites. En effet, les photos ou vidéos présentes dans les tweets n'ont pas été étudiées par souci d'automatisation. Il serait donc bon de se pencher un peu plus sur cet aspect pour trouver une solution car ces données peuvent être très riches et apporter un nouvel élément d'analyse novateur à notre étude.

Illustrations

FIGURE 01 – Diagramme représentant les bases de données	10
FIGURE 02 – Nombre de tweets par marques dans « <i>twitter_communauté</i> » avant et après le nettoyage des données	12
FIGURE 03 – Diagramme représentant le processus de collecte et de nettoyage des données	14
FIGURE 04 – Nombre de tweets et de « favorite » par page des marques.....	19
FIGURE 05 – Nombre de tweets et de « favorite » par marque pour tous les tweets et pour les tweets sans les réponses.....	21
TABLEAU 01 – Mots les plus récurrents et leur nombre d’occurrence par marque pour la base de données « <i>twitter_pages</i> ».....	24
TABLEAU 02 - Statistiques descriptives du nombre de retweet et de favorite.....	25
FIGURE 06 – Graphique à secteur représentant le pourcentage d’émotions dans le 1 ^{er} dataset	27
TABLEAU 03 – Nombre d’émotions par marque avec le 1 ^{er} ensemble de données	28
FIGURE 07 – Graphiques représentant la proportion d’émotions par marque avec le 1 ^{er} ensemble de données.....	29
FIGURE 08 - Graphique représentant la proportion d’émotions par marque avec le 2 ^{ème} ensemble de données.....	30
FIGURE 09 - Graphique représentant la proportion d’émotions par marque avec le 3 ^{ème} ensemble de données.....	30
FIGURE 10 – Graphique représentant le nombre de tweets en valeur absolue par marque et en fonction de la semaine	32
FIGURE 11 – Graphique représentant le nombre de tweets en pourcentage par marque et en fonction de la semaine	33
FIGURE 12 – Graphique représentant le nombre de tweets en pourcentage par marque (excepté Adidas) et en fonction de la semaine.....	33
FIGURE 13 – Graphique représentant le nombre de tweets en valeur absolue par marque et le nombre moyen de « favorite » en fonction du jour de la semaine	35
FIGURE 14 – Graphique représentant le nombre de tweets en pourcentage par marque et le nombre moyen de « favorite » en fonction du jour de la semaine	35
FIGURE 15 – Graphique représentant le nombre de tweets en valeur absolue par marque et le nombre moyen de « J’aime » en fonction de l’heure	37
FIGURE 16 – Graphique représentant le nombre de tweets en pourcentage par marque et le nombre moyen de « J’aime » en fonction de l’heure.....	37
TABLEAU 04 – Nombre de tweets en valeur absolue par catégorie et par marque	40
TABLEAU 05 – Pourcentage de tweets par catégorie et par marque	41
TABLEAU 06 – Mots les plus récurrents et leur nombre d’occurrence par marque pour la base de données « <i>twitter_communauté</i> ».....	45
FIGURE 17 – Graphiques représentant la proportion d’émotions par marque avec le 1 ^{er} ensemble de données.....	48
FIGURE 18 - Graphique représentant la proportion d’émotions par marque avec le 2 ^{ème} ensemble de données.....	49
FIGURE 19 - Graphique représentant la proportion d’émotions par marque avec le 3 ^{ème} ensemble de données.....	50

FIGURE 20 – Graphique représentant le nombre de tweets en pourcentage par communauté de marque et en fonction de la semaine	51
FIGURE 21 – Graphique représentant le nombre de tweets en pourcentage par communauté de marque (excepté Asics) et en fonction de la semaine	52
FIGURE 22 – Graphique représentant le nombre de tweets en pourcentage par communauté de marque et en fonction du jour de la semaine.....	54
FIGURE 23 – Graphique représentant le nombre de tweets en pourcentage par communauté de marque et en fonction de l’heure	55
FIGURE 24 – Graphique représentant le nombre de tweets et le chiffre d’affaire par marque.....	56

Bibliographie

- [1] Hervé, L. (2019). Les 50 chiffres à connaître sur les médias sociaux en 2019. *Récupéré le, 2.*
- [2] Greenwood, S., Perrin, A., & Duggan, M. (2016). Social media update 2016. *Pew Research Center*, 11(2), 1-18.
- [3] Bermingham, A., & Smeaton, A. F. (2010, October). Classifying sentiment in microblogs: is brevity an advantage?. In *Proceedings of the 19th ACM international conference on Information and knowledge management* (pp. 1833-1836).
- [4] Roesslein, J. (2018). tweepy Documentation.
- [5] MongoDB Inc. (2018). PyMongo 3.9.0 documentation.
<https://api.mongodb.com/python/current/tutorial.html>, consulté le 11/01/2020
- [6] Chodorow, K. (2013). *MongoDB: the definitive guide: powerful and scalable data storage*. "O'Reilly Media, Inc."
- [7] Crockford, D. (2012). Json. *ECMA International*.
- [8] Mottl D. (2019). GetOldTweets3. <https://pypi.org/project/GetOldTweets3/>, consulté le 11/01/2020
- [9] Bird, S., Klein, E., & Loper, E. (2009). *Natural language processing with Python: analyzing text with the natural language toolkit*. "O'Reilly Media, Inc."
- [10] Taylor, R. E. (1999). A six-segment message strategy wheel. *Journal of Advertising Research*, 39(6), 7-7.
- [11] Golan, G. J., & Zaidner, L. (2008). Creative strategies in viral advertising: An application of Taylor's six-segment message strategy wheel. *Journal of Computer-Mediated Communication*, 13(4), 959-972.
- [12] Jensen, M. B., & Jepsen, A. L. (2008). Online marketing communications: need for a new typology for IMC?. *Journal of Website promotion*, 2(1-2), 19-35.
- [13] Kim, A. J., & Ko, E. (2012). Do social media marketing activities enhance customer equity? An empirical study of luxury fashion brand. *Journal of Business research*, 65(10), 1480-1486.
- [14] Davis, R., Piven, I., & Breazeale, M. (2014). Conceptualizing the brand in social media community: The five sources model. *Journal of Retailing and consumer services*, 21(4), 468-481.

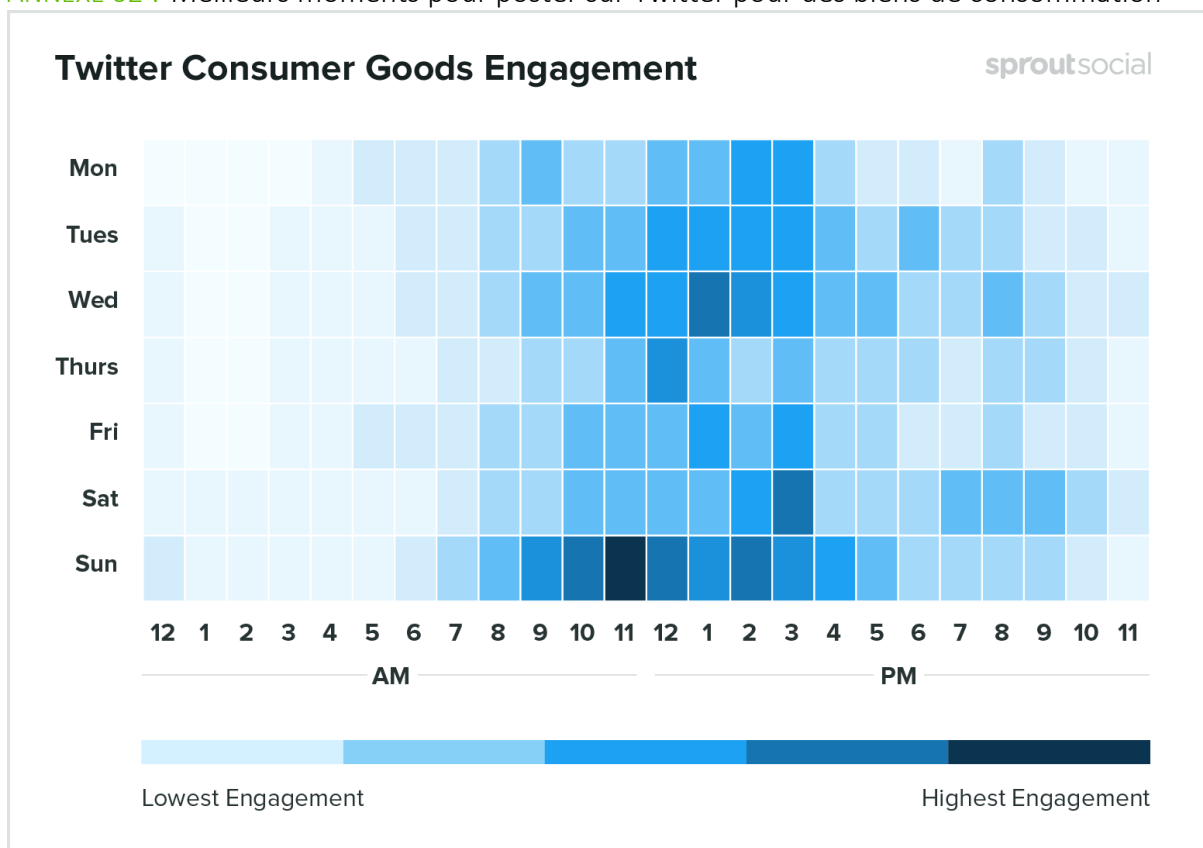
- [15] Coursaris, C. K., Van Osch, W., & Balogh, B. A. (2013, June). A Social Media Marketing Typology: Classifying Brand Facebook Page Messages For Strategic Consumer Engagement. In ECIS (p. 46).
- [16] Ashley, C., & Tuten, T. (2015). Creative strategies in social media marketing: An exploratory study of branded social content and consumer engagement. *Psychology & Marketing*, 32(1), 15-27.
- [17] Troussas, C., Virvou, M., Espinosa, K. J., Llaguno, K., & Caro, J. (2013, July). Sentiment analysis of Facebook statuses using Naive Bayes classifier for language learning. In *IISA 2013* (pp. 1-6). IEEE.
- [18] Boiy, E., Hens, P., Deschacht, K., & Moens, M. F. (2007, June). Automatic Sentiment Analysis in On-line Text. In *ELPUB* (pp. 349-360).
- [19] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Vanderplas, J. (2011). Scikit-learn: Machine learning in Python. *The Journal of machine Learning research*, 12, 2825-2830.
- [20] Chaffar, S., & Inkpen, D. (2011, May). Using a heterogeneous dataset for emotion analysis in text. In *Canadian conference on artificial intelligence* (pp. 62-67). Springer, Berlin, Heidelberg.
- [21] Novielli, N., Calefato, F., & Lanubile, F. (2018, May). A gold standard for emotion annotation in stack overflow. In *2018 IEEE/ACM 15th International Conference on Mining Software Repositories (MSR)* (pp. 14-17). IEEE. → emotions_dataset
- [22] Wang, B., Tsakalidis, A., Liakata, M., Zubiaga, A., Procter, R., & Jensen, E. (2016). SMILE Twitter Emotion dataset. *figshare*. <https://doi.org/10.6084/m9.figshare.3187909.v2>
- [23] Ekman, P. (1992). An argument for basic emotions. *Cognition & emotion*, 6(3-4), 169-200.
- [24] Arens, E. (2019). Best times to post on social media for 2019. *Verkkokauppa*. *Sprout Social*<<https://sproutsocial.com/insights/best-times-to-post-on-social-media/>>. Luettu, 6, 2019.
- [25] Spasojevic, N., Li, Z., Rao, A., & Bhattacharyya, P. (2015, August). When-to-post on social networks. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 2127-2136).
- [26] Hajikhani, A., Porras, J., & Melkas, H. (2017). Brand analysis in social network services: results from content analysis in twitter regarding the US smartphone market. *International Journal of Innovation and Technology Management*, 14(02), 1740008.

Annexes

ANNEXE 01 : Liste des mots à supprimer

```
{'before', 'it's', 'through', 'not', 'here', 'by', 'had', 'hasn', 'it', 'do', 'doesn't', 'wasn't', 'weren't', 'you'll', 'your s', 'again', 'himself', 'hadn't', 'further', 'who', 'haven', 'has', 'theirs', 'after', 'from', 'yourself', 'other', 'have', 'my self', 'won't', 'each', 'all', 'to', 'ours', 'when', 'your', 'her', 't', 'for', 'me', 'too', 'those', 'you've', 'themselves', 'him', 'and', 'until', 'only', 'you'd', 'whom', 'ma', 'mustn', 'shouldn', 'up', 'nor', 'this', 'no', 'own', 'hasn't', 'weren', 'did', 'both', 'she', 'against', 'because', 'under', 'which', 'about', 'our', 'ourselves', 'out', 'at', 'why', 'shouldn't', 'on ce', 'down', 'am', 'is', 'shan', 'more', 'if', 'between', 'you', 'should', 'we', 'these', 'was', 'just', 'aren', 'that'll', 'wh ere', 'during', 'any', 'wouldn't', 'm', 's', 'my', 'o', 'haven't', 'does', 'his', 'the', 'should've', 'shan't', 'don't', 'does n', 'with', 'then', 'same', 'you're', 'be', 'd', 'hers', 'into', 'few', 'below', 'ain', 'needn', 'as', 'doing', 'mightn't', 'su ch', 'll', 'an', 'most', 'while', 'on', 'off', 'can', 'don', 'mightn', 'isn't', 'she's', 'but', 'i', 'a', 'having', 'itself', 'on', 'needn't', 'their', 'mustn't', 'they', 'won', 'what', 'are', 'that', 'so', 'now', 'in', 'how', 'wouldn', 'couldn', 'ver y', 'y', 'there', 'than', 'aren't', 're', 'couldn't', 'of', 'didn', 've', 'didn't', 'some', 'above', 'its', 'he', 'them', 'had n', 'wasn', 'been', 'will', 'were', 'over', 'isn', 'yourselves', 'being', 'herself'}
```

ANNEXE 02 : Meilleurs moments pour poster sur Twitter pour des biens de consommation



ANNEXE 03 : Listes des championnats mondiaux parmi les plus grands événements sportifs de l'année 2019⁸

- 10 au 27 janvier - Handball : championnat du monde messieurs en Allemagne et au Danemark
- 14 au 27 janvier - Tennis/Grand Chelem : Open d'Australie à Melbourne
- 01 février - Football : finale de la Coupe d'Asie des nations à Abou Dhabi
- 01 février au 16 mars - Rugby : Tournoi des six nations : France - pays de Galles, Ecosse - Italie, Irlande - Angleterre
- 03 février - Football américain : Superbowl à Atlanta (Etats-Unis)
- 05 au 17 février- Ski alpin : championnats du monde à Are (Suède)

⁸ <https://sport24.lefigaro.fr/autres-sports/actualites/calendrier-sport-2019-939242>

- 19 février au 03 mars - Ski nordique : championnats du monde à Seefeld (Autriche)
- 21 février au 25 mars – CrossFit Games Open
- 17 mars - Auto/Formule 1: Grand Prix d'Australie (Melbourne)
- 20 mars au 29 septembre Baseball : Ligue Majeur de Baseball en Amérique du Nord
- 13 avril - Basket-ball/NBA : début des play-offs
- 14 avril - Cyclisme : Paris-Roubaix
- 15 avril – Marathon de Boston
- 26 mai au 09 juin - Tennis/Grand Chelem : tournoi de Roland-Garros à Paris
- 31 mai au 02 juin - Rugby : barrages de Top 14
- 01 juin - Football : finale de la Ligue des champions messieurs à Madrid
- 02 juin au 26 octobre – Course à pied : Golden Trail World Series
- 07 au 30 juin - Football : Coupe d'Afrique des nations
- 07 juin au 07 juillet - Football : Coupe du monde dames en France
- 17 juin– Salomon WMN camp
- 01 au 14 juillet - Tennis/Grand Chelem : tournoi de Wimbledon à Londres
- 06 au 28 juillet - Cyclisme : Tour de France
- 20 juillet – Tahoe Rim Trail
- 25 au 28 juillet – Athlétisme : championnats des Etats-Unis
- 25 aout - Pikes Peak Marathon
- 26 aout au 08 septembre - Tennis/Grand Chelem : US Open à New York (Etats-Unis)
- 31 aout au 15 septembre - Basket-ball : championnat du monde messieurs en Chine
- 20 septembre au 02 novembre - Rugby : Coupe du monde messieurs (Japon)
- 19 au 24 novembre - Phase finale de Coupe Davis dans sa nouvelle version (Madrid)
- 30 novembre au 15 décembre - Handball : championnat du monde dames au Japon
- 01 décembre - Auto/Formule 1 : Grand Prix d'Abu Dhabi (Yas Marina)

ANNEXE 04 : Listes comprenant les mots qui se rapportent à une stratégie de communication

```

social = ["earth", "save", "donate", "charity", "fundraiser", "nonprofit", "causes", "children", "hospital", "sustainable",
"social", "healthcare", "plastic", "recycle", "recycled", "recycling", "recyclable", "respect", "responsibly",
"responsible", "equal", "natgeo", "lisawells00"]
service = ["dm", "feedback", "local", "store", "service", "return", "suggestion", "warranty", "support", "email", "retailer",
"stores", "globalservicepuma.com", "message", "contact", "information", "help", "opening", "location"]
engagement = ["advice", "like", "facebook", "youtube", "snapchat", "instagram", "watch", "look", "creatorsunite", "check",
"favorite", "stability", "training", "website", "comfort", "energy", "messenger", "newsletter", "history", "follow"]
promotion = ["offer", "free", "chance", "sale", "luck", "discount", "coupon", "promo", "deal", "gratis", "special", "chance",
"reward", "contest", "giveaway", "sweepstake", "prize", "bostonmarathon", "championsleague"]
season = ["winter", "congrats", "congratulations", "christmas", "valentine", "newyear", "spring", "autumn", "summer", "mlb",
"boston", "wimbledon", "tahoe", "usatfoutdoors", "pikes", "goldentrailseries"]
celebrities = ["mariataylor", "karliekloss", "djokernole", "robinanderson", "bazleydarius", "cocogauff", "usainbolt",
"lewishamilton", "imvkhohli", "ybncondade", "orousteing", "selenagomez", "caradelevingne", "jjwatt", "gigihadid",
"anzeesen", "kilianj", "rickeygates", "markarmstrong", "rorysouthworth", "mikedski", "codytownsend",
"joamoyas", "martyvalmassoi", "clairetillyer", "nicovuignier", "ryansandes", "livsansoz", "michellanne",
"enkithalassa", "jamesmforrest", "ialnbethune", "tiarevincent", "tinaemellie", "marcuspreedy", "raflatuni",
"jimkchin", "alexhonnold", "lindseyvonn", "jordanspieth", "justinverlander", "anthonyfjoshua", "stephencurry30",
"stephencurry30s", "dennisismithjr", "dkmetcalf14", "bryceharper3", "stanrey7", "johnnoregan777", "iamcardib",
"reebokvictoriabeckham", "khuranarohit89", "carlwinslow187", "lfc", "steftsitsipas", "leomessi",
"thesergiogarcia"]
adidas_products = ["adizero", "stan", "smith", "superstar", "gazelle", "nmd", "campus", "samba", "eqt", "continental80",
"pharrell", "adicolor", "yung", "series", "ultraboost", "alphabounce", "solar", "supernova", "ozweego",
"zx", "adilette", "3mc", "coast", "star", "questar", "cosmic", "ivy", "supercourt", "predator", "nite",
"jogger", "copa", "clima", "rdy", "essentials", "trefoil", "teamwear", "parley", "primeblue", "graphics",
"designed4training", "nemeziz", "x", "primeknit", "alphaskin", "heat rdy", "adicross", "adipure",
"codechaos", "tour360", "terrex", "y3", "shoe", "shoes", "bra", "alphaedge"]
asics_products = ["courtffnovak", "risebryte", "flytefoam", "gelcumulus", "gelkayano", "metaride", "glideride", "evoride",
"gelnimbus", "gelexcite", "metarun", "gt2000", "gt1000", "gt4000", "gelexalt", "metaracer", "novablast",
"dynaflyte", "gel-ds trainer", "gel-noosa tri", "roadhawk", "gellyte", "gel-kinsei og", "gelquantum",
"gel1090", "gelnandi", "gelcontend", "future metallic", "gt series", "gelpulse", "gelfujitrabuco",
"gelresolution", "tartheredge", "gelsaga", "gelcitrek", "gelventure", "courtff", "geldedicate",
"gelchallenger", "solution speed", "gelgame", "court", "geltask", "sky elite", "netburner", "gelrocket",
"gelbeyond", "lethal", "menace", "gelbela", "gelpadel", "gelhunter", "gelblade", "upcourt", "fff",
"gelcumulus", "shoe", "shoes", "bra"]
newBalance_products = ["574", "997", "373", "996", "fresh foam", "990", "880", "860", "420", "x90", "247", "1500", "890",
"summit", "minimus", "991", "998", "530", "527", "sweat", "pullover", "jacket", "hoodie", "t-shirt",
"top", "trouser", "tight", "short", "pant", "bra", "omni1s", "furon", "990v5", "shoe", "shoes", "bra"]
nike_products = ["air", "pegasus", "zoom", "flyease", "flyknit", "free", "flywire", "lunarlon", "metcon", "sb", "jordan",
"superrep", "airforce", "blazer", "airmax", "vapormax", "huarache", "cortez", "presto", "joyride", "react",
"vomero", "structure", "kyrieirving", "kevindurant", "lebronjames", "phantom", "mercurial", "tiempo",
"sportswear", "swoosh", "blazer", "ryz", "court", "heritage", "foam", "shoe", "shoes", "bra"]
puma_products = ["cali", "rider", "pride", "rs", "rise", "deva", "firstmile", "xbionic", "lqd", "future", "one", "king",
"spark", "provoke", "kyron", "smash", "tazon", "enzo", "netfit", "nrgy", "lqdccl", "ignite", "carson",
"axelion", "hybrid", "pirelli", "nova", "cali", "central", "carina", "caracal", "stitch", "nucleus", "rsx",
"tmc", "cloud9", "merha", "shoe", "shoes", "bra"]
reebok_products = ["nano", "flexagon", "kinetica", "reago", "legacy", "elusion", "lite", "clubc", "aztrek", "classic",
"royal", "pheedan", "runner", "daytona", "instapump", "sublite", "energylux", "trail cruiser", "hiit",
"fitness epic", "solefury", "fury", "reebokclassics", "zig", "bra", "splitfrom", "floatride", "shoe",
"shoes", "bra"]
salomon_products = ["predict", "sonic", "sense ride", "vectur", "s/lab", "sense", "ultra", "cross", "speedcross", "wildcross",
"supercross", "alphacross", "alpine", "xa", "rx", "gtx", "vaya", "odyssey", "quest", "out", "outward",
"outback", "outline", "outbound", "raise", "reveal", "amphib", "crossamphibian", "shelter", "ellipse",
"outsnap", "outblast", "tech", "xt", "lightning", "sense", "elevate", "comet", "hoodie", "outspeed",
"grid", "icecrystal", "icetown", "icehearty", "icerocket", "icepuff", "qst", "brilliant", "fantasy",
"stormrace", "stormcozy", "stormgloss", "stormfluff", "stormpunch", "whitenight", "softshell", "agile",
"discovery", "primo", "wayfarer", "bonatti", "icefancy", "icemania", "stormseason", "moveon", "support",
"cross", "evasion", "nso", "xt6", "shoe", "shoes", "bra"]
theNorthFace_products = ["duffel", "backpack", "jacket", "parka", "anorak", "vest", "gilet", "coat", "hoodie", "cape",
"fleece", "pullover", "t-shirt", "bra", "top", "trousers", "jacket", "leggings", "suit", "shorts",
"pants", "shoes", "boots", "socks", "gloves", "hat", "beanie", "cape", "scarf", "bag", "visor",
"canister", "summit", "steep", "flight", "black", "shoe", "shoes", "bra"]
underArmour_products = ["hovr", "valsetz", "sonic", "tribase", "project rock", "charged", "skylar", "remix", "ripple",
"forge", "sleeve", "stellar", "jungle", "jacket", "shirt", "short", "polo", "hoodie", "heatgear",
"coldgear", "tank", "tech", "zip", "singlet", "vest", "short", "trouser", "jogger", "legging",
"tight", "underwear", "backpack", "bag", "headwear", "gloves", "socks", "belt", "capti", "phantom",
"uahovr", "curry7", "curry6", "uaforge96", "harper4", "shoe", "shoes", "bra"]
brands = ["adidas", "asics", "nb", "newbalance", "balance", "nike", "puma", "reebok", "salomon", "thenorthface", "tnf", "face
underarmour", "ua", "teamua"]

```